# Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments

Ning Wang[1], Ya Yang[2], Michael J. Moore[3], Samuel F. Brockington[4], Joseph F. Walker[1], Joseph W Brown[5], Bin Liang[1], Tao Feng[4], Caroline Edwards[3], Jessica Mikenas[3], Julia Olivieri[3], Vera Hutchison[3], Alfonso Timoneda[4], Tommy Stoughton[6], Raúl Puente[7], Lucas Majure[7], Urs Eggli[8], and Stephen A. Smith[1]

[1]Department of Ecology & Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048, USA

[2]Department of Plant and Microbial Biology, University of Minnesota-Twin Cities. 1445 Gortner Avenue, St. Paul, MN 55108 USA

[3]Department of Biology, Oberlin College, Science Center K111, 119 Woodland St., Oberlin, Ohio 44074 USA

[4]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[5]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

[6]Center for the Environment, MSC 63, Plymouth State University, 17 High Street Plymouth, NH 03264 USA

[7]Department of Research, Conservation and Collections, Desert Botanical Garden, 1201 N. Galvin Pkwy, Phoenix, AZ 85008 USA

[8]Sukkulenten-Sammlung Zürich, Mythenquai 88, CH-8002 Zürich, Switzerland

Corresponding authors: Ning Wang, Email: ningwum@umich.edu

Stephen A. Smith, E-mail: eebsmith@umich.edu

Abstract:
Survival in harsh environments is associated with several adaptations in plants. Species in the Portulacineae (Caryophyllales) have adapted to some of the most extreme terrestrial conditions on Earth, including extreme heat, cold, and salinity. Here, we generated 52 new transcriptomes and combined these with 30 previously generated transcriptomes, forming a dataset containing 68 species of Portulacinaeae, seven from its sister clade Molluginaceae, and seven outgroups. We performed a phylotranscriptomic analysis to examine patterns of molecular evolution within the Portulacineae. Our inferred species tree topology was largely congruent with previous analyses. We also identified several nodes that were characterized by excessive gene tree conflict and examined the potential influence of outlying genes. We identified gene duplications throughout the Portulacineae, and found corroborating evidence for previously identified paleopolyploidy events along with one newly identified event associated with the family Didiereaceae. Gene family expansion within Portulacineae was associated with genes previously identified as important for survival in extreme conditions, indicating possible molecular correlates of niche changes that should be explored further. Some of these genes also showed some evidence for positive selection. The correlation between gene function and expansion suggests that gene/genome duplication have likely contributed to the extreme adaptations seen in the Portulacineae.

key words: gene tree conflict, gene duplication, molecular evolution, paleopolyploidy, Portulacineae, stress adaptation

48 **Introduction**

49 Temperature and water availability are two major ecological variables that influence plant distribution
50 and survival (e.g., Peel et al. 2007). Plants living in severe environments that are characterized by
51 extremely low precipitation, extreme temperature ranges, intense sun and/or dry winds, often exhibit
52 morphological, physiological, and ecological characteristics that allow them to cope with such highly
53 variable abiotic stresses (Gibson 1996; Kreps et al. 2002). We may expect to find some signature of the
54 transitions in the phenotypic adaptations to survive these environments within genes and genomes.
55 While researchers have addressed these questions in several ways in different clades (e.g., Christin et al.
56 2007), until recently it has not been possible to obtain large scale genomic and transcriptomic data for
57 large clades.

58 The Portulacineae comprises ~ 2200 species in nine families and 160 genera (Nyffeler and
59 Eggli 2010a; Angiosperm Phylogeny Group 2016) that are inferred to have diverged from its sister
60 group ~54 Mya (Arakaki et al. 2011). Species in this suborder are incredibly morphologically diverse,
61 ranging from annual herbs to long-lived iconic stem succulents to trees. They also exhibit variable
62 habitat preferences ranging from rainforests to deserts (Smith et al. 2018). While some species exhibit a
63 worldwide distribution, the majority are restricted to seasonally dry areas of the northern and southern
64 hemisphere, under either hot arid or cold arid conditions (Hernández-Ledesma et al. 2015). Many
65 specialized traits such as fleshy or succulent stems, leaves, and/or underground parts, have arisen in
66 association with the adaptation of this clade to xeric, alpine and arctic environments. Portulacineae also
67 includes several transitions to Crassulacean acid metabolism (CAM) and $C_4$ photosynthesis that
68 improve photosynthetic efficiency and hence minimize water loss in hot and dry environments
69 compared to $C_3$ photosynthesis (Edwards and Walker 1983; Borland et al. 2009; Edwards and Ogburn
70 2012).

71 The Cactaceae encompasses roughly 80% of species in Portulacineae (~1850 species; Nyffeler
72 and Eggli 2010b) including one of the most spectacular New World radiations of succulent plants
73 (Mauseth 2006; Arakaki et al. 2011). In addition to succulent structures that enable water storage, they
74 exhibit an array of adaptations to cope with arid and semiarid conditions. This includes spines
75 (modified leaves) with some advantageous effects (Anderson 2001) and extensive, but relatively
76 shallow, root systems that facilitate quick absorption of rainfall (Gibson and Nobel 1990). In addition
77 to Cactaceae, Didiereaceae exhibit similar but less pronounced adaptations to the hot and seasonally
78 dry environments of Madagascar (Arakaki et al. 2011). By including only 20 species, Didiereaceae
79 have lower species diversity compared to Cactaceae, whereas they are no less interesting in their
80 specialized adaptations, including the evolution of thorns and succulent leaves. In contrast to the

81 largely tropical and subtropical Cactaceae and Didiereaceae, Montiaceae has a cosmopolitan
82 distribution with most species occurring in colder environments. Some even inhabit high alpine zones
83 and/or the high Arctic, especially for species within the genera *Claytonia* and *Montia* (Ogburn and
84 Edwards 2015; Stoughton et al. 2017). The repeated adaptations to harsh environments and remarkable
85 habitat disparity within the Portulacineae suggest the potential that ancestral genomic elements may be
86 predisposed to repeated co-option within each clade.

87 In the face of rapid environmental changes, understanding the capacity of plants to survive and
88 evolve in different environments requires study of the genetic basis of adaptation to abiotic stresses
89 (e.g., cold, hot, and dry conditions; Chaves et al. 2003; Fournier-Level et al. 2011). For example, genes
90 involved in the abscisic acid (ABA) signaling pathway [e.g., ABRE-binding protein/ABRE-binding
91 factors, the dehydration-responsive element binding (DREB) factors, and the NAC transcription factors
92 (e.g., Qin et al. 2004, Nakashima et al. 2012)] have been demonstrated to function in drought, cold, and
93 heat responses (Nakashima et al. 2014). However, plant stress responses may be complicated by
94 multiple stresses (Nakashima et al. 2014), and so the role of genes to stress are still mostly unknown.
95 Thus, by examining genes and genomes in plant lineages that have experienced macroevolutionary
96 adaptations to stressful conditions we may yet gain new insights into the evolution of stress responses
97 that may be broadly applicable to plant biology. Given their propensity for high stress environments,
98 diverse ecophysiology, and relatively high number of species, the Portulacineae is an excellent group to
99 explore such patterns.

100 Recent phylogenetic work has helped resolve relationships within the Portulacineae, although
101 several key nodes around the early divergence of Portulacineae remain only moderately or poorly
102 supported (Arakaki et al. 2011; Ogburn and Edwards 2015; Yang et al. 2018; Moore et al. 2017;
103 Walker et al. 2018a). Several of these nodes may reflect rapid radiations (Arakaki et al. 2011) that
104 deserve further examination. In addition, recent studies have inferred multiple paleopolyploidy events
105 in Portulacineae, including one at the base of the clade (Yang et al. 2018). With increased taxon
106 sampling, we may detect additional whole genome duplication (polyploidy) events. In this study, we
107 analyzed 82 transcriptomes (of which 52 are newly generated) to better understand the evolutionary
108 history of Portulacineae. By closely examining gene-tree/species-tree conflict, patterns of gene
109 duplication and paleopolyploidy, and clade-specific evolutionary patterns of genes associated with
110 stress-related responses, we aimed to 1) identify nodes with significant gene tree discordance and
111 reveal the underlying mechanisms that are responsible for conflicting hypotheses; 2) confidently
112 identify lineage-specific genome duplication events through improved taxon sampling within

113 Portulacineae; and 3) test whether gene/genome duplications are associated with clade-specific
114 adaptive traits.
115
116 **Results and Discussion**
117
118 **Phylogeny of Portulacineae**
119 Our concatenated supermatrix contained 841 gene regions and had a total aligned length of 1,239,871
120 bp. The matrix had a gene occupancy of 95.4% and character occupancy of 84.5%. Of the ingroup taxa,
121 only *Pereskia aculeata* had a gene occupancy less than 80%, while the majority (67 taxa) had over 90%
122 of genes present in the final supermatrix.
123 Both the concatenated maximum likelihood tree (hereafter the CML tree) and the maximum
124 quartet support species tree (MQSST) recovered the same topology with similar support values (fig. 1),
125 all of which were largely consistent with previous estimates of the Portulacineae phylogeny. However,
126 because bootstraps may be a poor indicator of support in large phylogenomic datasets (e.g., wrong
127 phylogeny can be increasingly supported as more sequence data are added, Alfaro et al. 2003; Phillips
128 et al. 2004), we also conducted gene tree conflict analyses (fig. 2A, Smith SA et al. 2015). For
129 example, we found strong support (100% multi-locus bootstrap from ASTRAL and ML bootstrap) for
130 the sister relationship of Molluginaceae and Portulacineae (fig. 1), consistent with several recent
131 studies (Edwards and Ogburn 2012; Yang et al. 2015, 2018; Moore et al. 2017), but also found that
132 most gene trees were either uninformative for this clade or conflicted with the ML result. Our analysis
133 recovered the ACPT clade (Anacampserotaceae, Cactaceae, Portulacaceae and Talinaceae; Hershkovitz
134 and Zimmer 1997; Applequist and Wallace 2001; Applequist et al. 2006; Nyffeler 2007; Nyffeler and
135 Eggli 2010b; Ocampo and Columbus 2010; Arakaki et al. 2011; Moore et al. 2017), with high support
136 across genes trees for a topology of (((A, P), C), T) (figs. 1-2). The resolution of major clades within
137 Portulacineae generally agrees with that of Moore et al. (2017) using a targeted enrichment approach.
138 Our broad taxon sampling also allowed for resolution of many deeper relationships within
139 Portulacineae families. We found support for the sister relationship of Didiereaceae subfamilies
140 Portulacarioideae (*Ceraria* and *Portulacaria*) and Didiereoideae (Madagascar Didiereaceae;
141 Applequist and Wallace 2003; fig. 1), and although many gene trees were uninformative at this node,
142 the ML resolution was the dominant gene tree topology (fig. 2A). While relationships among the
143 genera within the Didiereoideae have been difficult to resolve with targeted-gene analyses (Applequist
144 and Wallace 2000; Nyffeler and Eggli 2010b), we recovered very strong support (100%) in gene tree
145 analyses for a clade of ((*Alluaudia*+*Alluaudiopsis*), (*Didierea*+*Decarya*)). This result differs from

146 Bruyns et al. (2014) that found *Alluaudiopsis* to be sister to the remaining Didiereoideae.
147 Within Cactaceae, we recovered three major clades (*Pereskia* s.s., *Leuenbergeria,* and the core
148 cacti, as in Edwards et al. 2005, Bárcenas et al. 2011) with strong support (> 97%, fig. 1), but with
149 substantial gene tree discordance (see below, fig. 2A). In fact, the majority of genes (>85%) in the
150 earliest diverging nodes within Cactaceae were either uninformative or conflicted with each other (fig.
151 2A). Within the core cacti, we recovered *Maihuenia* as sister to Cactoideae with strong support, and
152 this clade as sister to Opuntioideae as found by Edwards et al. (2005) and Moore et al. (2017).
153 Nevertheless, the position of *Maihuenia* has previously been found to be highly unstable (e.g., within
154 Opuntioideae, Butterworth and Wallace 2005; sister to Opuntioideae, Nyffeler 2002; or sister to
155 Opuntioideae+Cactoideae, Hernández-Hernández et al. 2014; Moore et al. 2017). Unlike other areas
156 within the Cactaceae that have high gene tree discordance, the monophyly of Opuntioideae (e.g.,
157 Barthlott and Hunt 1993; Wallace and Dickie 2002; Griffith and Porter 2009; Hernández-Hernández et
158 al. 2011) and Cactoideae (Nyffeler 2002; Bárcenas et al. 2011) had relatively high concordance (95-97%
159 gene trees support these monophyletic groups; fig. 2A). The high level of discordance among gene
160 trees within Opuntioideae likely reflects the difficulty of resolving relationships among tribes that
161 could not be monophyletic (e.g., Cylindropuntieae, Opuntieae; Edward et al. 2005; Bárcenas et al. 2011;
162 Hernández-Hernández et al. 2011). However, these conflicts may also be attributed to limited taxon
163 sampling as we only have one or two species representing each tribe (figs. 1 and 2A). The topology of
164 the Cactoideae was largely congruent with previous studies (e.g., Butterworth et al. 2002; Nyffeler
165 2002; Bárcenas et al. 2011; Hernández-Hernández et al. 2011). Within Cacteae, the resolution of the
166 five sampled genera was strongly supported and are consistent with what was recovered by Hernández-
167 Hernández et al. (2011) and Vázquez-Sánchez et al. (2013) using five loci. However, the core
168 Cactoideae (sister to Cacteae, Hernández-Hernández et al. 2011) was poorly supported (52% CML and
169 54.5% MQSST, fig. 1) in both species tree inferences. With limited taxon sampling, some relationships
170 (e.g., the monophyly of the Pachycereeae) are largely consistent with previous classification (e.g.,
171 Nyffeler and Eggli 2010b), whereas others (e.g., the position of *Gymnocalycium*) are still controversial
172 among studies (Arakaki et al. 2011; Bárcenas et al. 2011; Hernández-Hernández et al. 2014).
173
174 **Assessing conflict on specific nodes**
175 As genomes and transcriptomes have become increasingly available, our ability to analyze
176 phylogenetic conflict and concordance across hundreds or thousands of genes has increased
177 significantly. One of the major findings of these genome-scale studies is that gene tree conflicts are
178 prevalent throughout the tree of life (e.g. Degnan and Rosenberg 2009, Jarvis et al. 2014).

179 Incongruence among genealogies can result from numerous processes, including incomplete lineage
180 sorting (ILS), horizontal gene transfer (HGT), and gene duplication and loss (Maddison, 1997).
181 Methods that explicitly model the source of discordance among multi-locus data are very important for
182 correctly inferring the underlying species tree (Liu and Pearl 2007; Liu et al. 2008), but a close
183 examination of gene-specific discordance is also valuable because it may reveal the underlying
184 processes that characterize the evolutionary history of a given clade (e.g., Shen et al. 2017). Above, we
185 discussed gene tree conflict as it pertains to support, or lack thereof, for different clades. To specifically
186 address gene-specific support for alternative relationships, we examined individual genes and their
187 support for nodes, focusing on phylogenetic relationships that had previously been identified to have
188 significant conflicts. With these conflict assessments, we examined one focal node. This allowed for
189 alternative resolutions at a node while allowing for the remainder of the tree to be optimized for the
190 highest maximum likelihood topology. Therefore, even though two genes may support the same
191 resolution for the node of interest, they may have completely different topologies in areas excluding
192 this node. This focused analysis allowed us to isolate the signal at particular nodes without having to
193 accommodate the many diverse processes that may shape conflict in other parts of the tree.
194 Within Portulacineae, several regions of the tree have been identified in previous studies as
195 having significant conflict and/or low support, e.g. the early divergences of Cactaceae, the relationship
196 among Portulacaceae, Cactaceae, and Anacampserotaceae, and the positions of Basellaceae and
197 Didiereaceae (e.g., Moore et al. 2017, Walker et al. 2018a). For example, Portulacaceae has been
198 recovered as sister to Anacampserotaceae (Ogburn and Edwards 2015; Walker et al. 2018a; and
199 recovered here) and to Cactaceae (Arakaki et al. 2011; Moore et al. 2017). We examined the three
200 topological positions among these families to determine the gene-wise pattern of support for each
201 topology. Within the above three topological constraints, 476 (out of 841) genes supported
202 Anacampserotaceae+Portulacaceae, 164 genes supported Cactaceae+Portulacaceae, and 201 genes
203 supported Cactaceae+Anacampserotaceae.
204 Recently, studies have demonstrated that only a few genes in phylogenomic datasets can drive
205 the resolution of nodes (e.g., Brown and Thomson 2017; Shen et al. 2017; Walker et al. 2018b). Using
206 these node-specific analyses, we found several outlying genes (i.e., genes highly favor one topology
207 over the other) despite high topological support as measured by bootstraps (fig. 2B). For example, one
208 gene (cluster4488, with homology to the "ARABILLO 1-like" genes in *Beta vulgaris*) supported
209 Cactaceae+Portulacaceae strongly over the others (>60 ln*L* units). Another gene (cluster7144, with
210 homology to the "UV-B induced protein chloroplastic-like" gene in *Chenopodium quinoa*) supported

211 Cactaceae+Anacampserotaceae strongly (>100 ln*L* units over the other relationships). Neither of these
212 genes exhibited obvious alignment errors.
213 We conducted similar analyses for the resolution of Basellaceae and Didiereaceae and for the
214 resolutions of the early diverging relationships within the Cactaceae. In specific, 506 genes supported
215 (Basellaceae, (Didiereaceae+others)) and 334 genes supported Basellaceae+Didiereaceae as found by
216 Soltis et al. (2011) and Anton et al. (2014). No significant outlying genes were observed supporting
217 either the above topological constraints (fig. 2B). For the early diverging Cactaceae, we examined the
218 genes that supported the grade of relationships (*Leuenbergeria,* (*Pereskia*+others)) versus
219 *Leuenbergeria*+*Pereskia*. Overall, 472 genes supported the grade, with 368 supporting
220 *Leuenbergeria*+*Pereskia*. We also found one gene (cluster4707, with homology to the "NF-X1-type
221 zinc finger protein") strongly supported the grade topology (>110 ln*L* units), without any obvious
222 mismatching in alignment.
223
224 **WGD featured the evolution of Portulacineae**
225 Whole-genome duplications (WGD) have been recognized as having a profound influence on the
226 evolutionary history of extant lineages, especially for plants (e.g., Cui et al. 2006; Soltis and Soltis
227 2009; Jiao et al. 2011; Wendel 2015; Yang et al. 2015; Smith et al. 2018). Major WGD events have
228 been inferred within the Caryophyllales (Yang et al. 2015, 2018; Walker et al. 2017) including three
229 within Portulacineae: in the ancestor of Portulacineae, in the ancestor of Basellaceae, and within
230 Montiaceae. The improved taxon sampling in the current study enabled us to infer additional events of
231 putative WGD and large-scale gene duplications (fig. 1). Specifically, large-scale gene duplications
232 were found in the ancestor of Portulacineae (node 13), subclades of Montiaceae [along the origin of
233 *Calyptridium* (node 17), and the sister clade of *Calandrinia* (node 19)], in the ancestor of Didiereoideae
234 (node 27), and in the ancestor of Cactaceae, Cactoideae and Opuntioideae (nodes 35, 39 and 50).
235 We considered evidence for a WGD event in the ancestor of Portulacineae (Yang et al. 2015) to
236 be a high percentage of gene duplications (13.4%) and a Ks peak (~0.4-1.0) shared by all members in
237 this clade. Our identification of putative WGD at the base of Portulacineae contrasts with the analyses
238 of Yang et al. (2018) that found increased gene duplication (24%) in the ancestor of
239 Portulacineae+Molluginaceae. We did not detect a similar pattern of gene duplications at this node (1%,
240 fig. 1), confirming that the Yang et al. (2018)'s estimation could result from phylogenetic uncertainty.
241 We found a high percentage of gene duplication for the Didiereoideae (50.1%, fig. 1) accompanied by
242 a very recent Ks peak (between 0.05-0.10) for all the species within this clade. We also identified
243 genome duplications within the Montiaceae. Yang et al. (2018) identified WGD in the ancestor of

244 *Claytonia* species. With expanded sampling, we placed this WGD event at the ancestor of *Lewisia* and
245 *Claytonia*, as all four species in this clade exhibited Ks peak between 0.2-0.4 (supplementary figs. S2
246 and S3, Supplementary Material online). We also identified an elevated number of gene duplications in
247 the two *Calyptridium* species, accompanied by a very recent Ks peak (between 0.05-0.10; fig. 1).
248 While we identified a high number of gene duplications at the origin of the Cactaceae, we did
249 not infer a WGD event as we did not detect a corresponding elevated Ks peak. This finding is
250 consistent with other recent studies (Walker et al. 2017; Yang et al. 2018). Several other nodes within
251 the phylogeny (e.g., nodes 50 and 39) exhibited similar patterns with high numbers of duplications but
252 no corresponding Ks peak (fig. 1). There are several complications (gene loss, gene tree conflict, life
253 history shifts, etc.) that may obscure the detection of gene/genome duplications and further
254 investigation with more taxa are likely to help shed light in future studies.
255 For identifying duplications within a single lineage, we examined Ks plots as gene tree-based
256 methods cannot accurately infer WGD on terminal branches. Our Ks analyses inferred two additional
257 genome duplication events occurring within single taxa: in *Basella alba* (Ks=0.45) and *Mollugo*
258 *verticillata* (Ks=0.25) (supplementary fig. S2, Supplementary Material online). Both species showed
259 higher chromosome counts relative to close allies (Yang et al. 2018). These WGD inferences are
260 consistent with previous studies (Yang et al. 2015, 2018).
261 Although Ks plots have been widely applied to WGD inference, higher Ks values (e.g., >0.75)
262 are associated with increasingly large error (Li 1997), so it is more appropriate to detect recent WGDs
263 with Ks values less than 2 (Vanneste et al. 2013). On the other hand, peaks at smaller Ks values (e.g., <
264 0.25) can also be difficult to interpret due to the splice variants or other anomalies produced during
265 transcriptomic assembly. However, in the case of slowly evolving lineages or recent duplications, only
266 examining larger Ks can result in missing of some potential duplications. We note that this may be the
267 case with the Didiereoideae where the Ks peak falls in a range that also may contain splice variants (i.e.,
268 Ks peak occurred < 0.1, supplementary fig. S2, Supplementary Material online).
269 Some studies have suggested that some WGDs may be associated with adaptations to extreme
270 environments (e.g., Stebbins 1971; Soltis and Soltis, 2000; Brochmann et al. 2004),
271 speciation/diversification (Stebbins 1971; Wood et al. 2009), and success at colonization of new
272 regions (Soltis and Soltis, 2000). Smith et al. (2018), using the information regarding duplications from
273 Yang et al. (2018), found WGDs within the Caryophyllales to be associated with shifts in climatic
274 niche. The addition of a WGD involving the Didiereoideae support this finding. Unlike their sister
275 clade (including *Ceraria pygmaea* and *Portulacaria afra*) that is mainly distributed in the mainland
276 southern Africa, all species in Didiereoideae live in the ecoregion of southwest Madagascar island.

277 While we do not suggest that WGDs are always associated with shifts in niche or ecological setting,
278 within the Caryophyllales, this appears to be a common pattern.
279
280 **Lineage-specific gene expansions associated with adaptive traits**
281 Analyses on Gene Ontology (GO) overrepresentation were conducted on genes duplicated in nodes 15,
282 17, 19, 26, 27, 35, 39, 50, and 71 (fig. 1). Of genes duplicated at each clade, 30%-55% had identifiable
283 *Arabidopsis* IDs and almost all could be mapped in PANTHER (table 1). Genes belonging to "calcium
284 ion binding" (GO:0005509) showed significant overrepresentation at the origins of the Montiaceae
285 (node 15; except for *Phemeranthus parviflorus*) and the Didiereoideae (node 27); genes belonging to
286 "sulfur compound metabolic process" (GO:0006790) also exhibited significant overrepresentation at
287 the origin of Didiereoideae; and GO overrepresentation in "Hydrolase activity, acting on ester bonds"
288 (i.e., esterase, GO: 0016788, $p = 0.028$, table 2) was found at the origin of the Cactaceae (Node 35).
289 Many of the enriched GO terms described functional classes that are involved in diverse processes,
290 such as "calcium ion binding" linked the universal secondary messenger calcium, or "hydrolase
291 activity" linked to ester bond chemistry. However, others may be plausibly linked to potential adaptive
292 traits in the Portulacineae. In this regard, the expansion of genes belonging to the "sulfur compound
293 metabolic process" (GO:0006790) are of interest because sulfur-bearing evaporate compounds
294 (principally gypsum) are commonly found in soils in areas with low rainfall and high evaporation rates
295 (Watson 1979), such as the hot and cold deserts in which many lineages of the Portulacineae have
296 diversified. Researchers have also connected primary sulfur metabolism (e.g., sulfate transport in the
297 vasculature, its assimilation in leaves, and the recycling of sulfur-containing compounds) with drought
298 stress responses (Chan et al. 2013). Specifically, sulfate implicated in ABA-induced stomatal closure
299 and drought-responsive metabolites have been shown to be coordinated with sulfur metabolic pathways
300 (Chan et al. 2013). Consequently, duplication and overrepresentation of these sulfur metabolic genes
301 could potentially be evidence of adaptation to corresponding harsh environment of hot and cold deserts
302 (e.g., Lee et al. 2011). However, while it is tempting to speculate that these duplications have been
303 maintained through adaptation, we recognize that such 'duplication' events may simply be the
304 remnants of ancient polyploidy. In the absence of functional characterization, it is also necessary to
305 interpret all GO term analyses with caution.
306
307 **Gene families show broad expansions across Portulacineae**
308 In addition to looking at lineage-specific gene expansions, we also examined gene-family expansion
309 without reference to a specific organismal lineage within Portulacineae. This same approach has been

310 instrumental in revealing gene expansion and neo-functionalization implicated in the evolution of
311 betalains (Yang et al. 2015; Brockington et al. 2015) in Caryophyllales. Here, we explored the top 20
312 most expanded gene families (hereafter, called TOP20, table 3). The TOP20 included genes encoding
313 transporters, proteases, cytoskeletal proteins and enzymes that involved in photosynthetic pathways
314 (table 3). Within them, some are implicated in responses to drought: the gene encoding plasma
315 membrane intrinsic protein (PIP) is an aquaporin that can regulate the transport of water flux through
316 membranes (Vandeleur et al. 2009), heat shock proteins can prevent proteins from aggregating or being
317 rendered nonfunctional (Kiang and Tsokos 1998), and ubiquitin can correct or degrade misfunctional
318 proteins (Hochstrasser 2009). But perhaps the most interesting gene radiations involve those encoding
319 Phosphoenol-pyruvate carboxylase (PEPC, Silvera et al. 2014) and NADP-dependent malic enzyme
320 catalyses (Ferreyra et al. 2003) that are key enzymes involved in CAM/$C_4$ photosynthesis (Smith and
321 Winter 1996; Cushman 2001). The role of gene duplication in the evolution of $C_4$ photosynthesis has
322 been contentious, and some authors have proposed that neo-functionalization of genes following
323 duplication has not played a major role in the evolution of $C_4$ photosynthesis (Gowik and Westhoff
324 2011; Williams et al. 2012; van den Bergh et al. 2014). However, recent analyses provide evidences
325 that duplication and subsequent retention of genes is associated with the evolution of $C_4$ photosynthesis
326 (Emms et al. 2016), specifically including the NADP-dependent malic enzyme. Moreover, convergent
327 evolution in several key amino acid residues of PEPC has been suggested to be associated with the
328 origin of both $C_4$ and CAM (Christin et al. 2007; Christin et al. 2014). A recent study has also
329 confirmed the occurrence of multiple rounds of duplication within the major PEPC paralog (PEPC1E1)
330 in the ancestral Portulacineae (Christin et al. 2014). Our analyses confirm that these genes have
331 experienced expansions within the Portulacineae, suggesting a role of this process in the evolution and
332 adaptation in this group.
333
334 **Targeted analyses of drought and cold associated genes**
335 We focused on a set of 29 functionally annotated genes known to be involved in drought and cold
336 tolerance (table 4) as the Portulacineae have repeatedly evolved into cold and drought environments.
337 We found almost all of these genes have experienced gene duplications including at the origin of
338 Diediereoideae, within Montiaceae and within Cactaceae when only including nodes with SH-like
339 support > 80 (fig. 3). Some of these lineages have experience WGDs, and so detecting gene-
340 duplications with these genes is, perhaps, unsurprising. However, their maintenance and, in some cases,
341 further expansion is notable. For example, we found duplications in WIN1 (SHN1) proteins in both
342 Diediereoideae and Montiaceae (table 4). These proteins belong to the Apetala2/Ethylene Response

343 Factor (AP2/ERF) family and are transcription factors associated with epicuticular wax biosynthesis
344 that increases leaf surface wax content. Overexpression of this gene can reduce direct water loss
345 through the cuticular layer and increase plant tolerance to abiotic stresses (Sajeevan et al. 2017).
346 It is also noteworthy that many of duplications in these 29 focal genes occurred at the origin of
347 Portulacineae (6 genes in table 4) or earlier (fig. 3). This suggests the possibility that some genes
348 associated with adaptations to harsh environments diversified early in the evolution of the
349 Portulacineae and the Caryophyllales. While different lineages may have adapted to different
350 environments throughout the evolution of the Portulacineae, their predisposition to become adapted to
351 these environments may be the result of the early diversification of the gene families. Additionally,
352 genes from the same family may experience different evolutionary histories. For example, six
353 homologs that are part of the CDPK (caldium-dependent protein kinase, also CPK) gene family are
354 known to be involved in drought stress regulation (Geiger et al. 2010; Brandt et al. 2012). However,
355 the number of duplications varied between different lineages, regardless of the similarity in habitat
356 [four duplicated at the origin of Didiereoideae, one within Montiaceae (node 17), one at Portulacaceae,
357 and two within the Cactaceae (nodes 35 and 57)].
358 These 29 targeted genes also exhibited variable lineage-specific positive selection within
359 Portulacineae (table 4). The strongest signals of positive selection were all found in genes associated
360 with drought and/or cold tolerance in the ABA signaling pathway (such as the NAC10/29 and
361 WRKY33 TFs, the CDPK18 and the PP2Cs, table 4, e.g., Golldack et al. 2014; Nakashima et al. 2014;
362 Huang et al. 2015; Li et al. 2017), a central regulator of abiotic stress resistance in plants. Interestingly,
363 the ABI1 and PP2CA gene (Mishra et al. 2006; Merlot et al. 2001), which encode two proteins of the
364 PP2Cs family, were among the genes with the highest number of lineages under positive selection in
365 our tests. As negative regulators, PP2Cs can regulate numerous ABA responses, such as stomatal
366 closure, osmotic water permeability of the plasma membrane, drought-induced resistance and
367 rhizogenesis, seed germination and cold acclimation (Gosti et al. 1999; Merlot et al. 2001; Mishra et al.
368 2006). In addition to PP2C genes, the positive regulators SNF1-related protein kinase 2s (SnRK2s) are
369 also related to the ABA signaling module (Hubbard et al. 2010). While we did not detect positive
370 selection in the three SnRK2s genes (i.e., SnRK2.4, 2.5, 2.6, table 4), we found that they experienced
371 ancient duplications at the origin of Portulacineae or earlier and recent duplications within Montiaceae,
372 Didiereaceae, or Cactaceae (nodes 17, 27, 39; table 4). Positive selection and gene family expansion in
373 some of these gene families within Cactaceae, Didieraceae, and Montiaceae suggests that these genes
374 warrant further investigation for their potential role in the evolution of adaptations to challenging
375 environmental conditions.

376

377 **Conclusion**

378 Using a transcriptomic dataset spanning across 82 species, we reconstructed a phylogeny of
379 Portulacineae that reveals significant underlying gene-tree/species-tree discordance. These conflicts
380 may arise from an array of biological sources (e.g., hybridization, ILS), and taxon sampling (even with
381 large number of genes, Walker et al. 2017) or sequence type may also influence species topologies in
382 phylogenomic analyses (Reddy et al. 2017). We examined specific nodes in detail to attempt to
383 document how much support there may be for particular relationships, regardless of conflict in the rest
384 of the tree. Our findings, along with others recently published (Smith SA et al. 2015; Brown and
385 Thomson 2017; Shen et al. 2017; Walker et al. 2018b), further demonstrate the need and importance to
386 examine gene tree relationships in detail. As genome and transcriptome studies continue to document
387 significant conflict, we need to consider different means of resolving relationships that can filter or
388 accommodate these signals. Furthermore, the conflicts within these data likely hold interesting
389 evolutionary questions and answers that should be analyzed in more detail.

390 With increased sampling, we also recorded new paleopolyploidy events. This contributes to a
391 growing body of literature that is documenting the importance of these events in the evolution of plants
392 (e.g., Cui et al. 2006; Smith et al. 2018). Also, as has been repeatedly demonstrated, with additional
393 sampling effort, we are likely to find additional duplication events. The importance of these
394 duplications for the evolution of individual clades needs to continue to be explored in detail. The gene
395 expansion and genome duplications found within the Portulacineae seem to be associated with shifts
396 into different environments and climates. This has been suggested by previous researchers (e.g., Smith
397 et al. 2018) and should be examined further across other diverse angiosperm lineages.

398 Finally, by examining gene families in detail we are able to better highlight potential molecular
399 evolutionary patterns associated with particular adaptations. However, the approaches explored here
400 and by other researchers have major limitations as they rely on annotations based on model organisms
401 that may be distantly related to the species of interest. While the gene may have interesting patterns, the
402 GO labeled to the gene may or may not carry much meaning. These genes may have different functions
403 in these different lineages and so the results from these analyses should not be over interpreted. Despite
404 these limitations, however, these analyses uncover interesting gene regions that should be explored
405 further, regardless of the accuracy of the label, as the patterns of selection and duplication alone
406 suggest their importance in the evolution of the lineage, even if not related to the particular adaptation
407 of interest. As angiosperms continue to be more thoroughly sequenced, we will continue to shed light
408 on the molecular evolutionary patterns and processes that have shaped the tree of life.

409

410 **Material and Methods**

411

412 **Taxon sampling and transcriptome generation**

413 Sixty-eight ingroup species were included in this study, representing all Portulacineae families
414 (Anacampserotaceae, Cactaceae, Basellaceae, Didiereaceae, Montiaceae, Talinaceae, and
415 Portulacaceae) *sensu* APG IV (Angiosperm Phylogeny Group, 2016) except for Halophytaceae
416 (supplementary tables S1 and S2, Supplementary Material online). We also included 14 Caryophyllales
417 species as outgroups, including seven taxa of Molluginaceae (supplementary tables S1 and S2,
418 Supplementary Material online). Of the 82 transcriptomes included in the study, 52 were newly
419 generated following the protocol of Yang et al. (2017), which is briefly described below
420 (supplementary table S1, Supplementary Material online). RNA was extracted from flash-frozen young
421 leaves and/or flower buds using either the Ambion PureLink Plant RNA Reagent (ThermoFisher
422 Scientific Inc, Waltham, MA, USA), Aurum™ Total RNA Mini Kit, or a hot acid phenol protocol
423 (Yang et al. 2017). Total RNA was quantified with the Agilent 2100 Bioanalyzer (Agilent
424 Technologies, Santa Clara, CA, USA). It was then used for library preparation using either the TruSeq
425 RNA Sample Preparation Kit v2 (Illumina, Inc., San Diego, CA) or the KAPA Stranded mRNA-Seq
426 kit (KAPA Biosystems, Wilmington, Massachusetts, USA), and sequenced on an Illumina HiSeq
427 platform at the University of Michigan DNA Sequencing Core (see deposited SRA, BioProject: XXX)
428 (supplementary table S1, Supplementary Material online).

429 On average, ~30 million paired-end reads were generated for each transcriptome and corrected
430 for error in Rcorrector (Song and Florea 2015). Corrected reads were assembled using Trinity v2.2.0
431 (Grabherr et al. 2011) after filtering of Illumina adapters (Illuminaclip 2:30:10) and low quality reads
432 (sliding window 4:5, leading 5, trailing 5 and min length 25) using Trimmomatic v0.36 (Bolger et al.
433 2014). Transdecoder v2.0 (Haas et al. 2013) was then used to translate assembled transcripts with
434 guidance of BLASTP against a concatenated proteome database including *Arabidopsis thaliana*
435 (TAIR10) and *Beta vulgaris* (Dohm et al. 2014, RefBeet-1.2.2)(release-34 from EnsemblPlants,
436 http://plants.ensembl.org/info/website/ftp/index.html). Translated amino acid sequences (PEP) and
437 coding DNA sequences (CDS) of the newly collected samples, as well as 30 publically available
438 datasets (Yang et al. 2015; Brockington et al. 2015; Walker et al. 2018a), were used for subsequent
439 analyses (supplementary table S2, Supplementary Material online). Highly similar sequences were
440 reduced using CD-HIT v4.6 for each dataset (PEP: -c 0.995 -n 5; CDS: -c 0.99 -n 10, Fu et al. 2012)

441

442 **Homology Inference**

443 Homology inference was carried out following the pipeline of Yang and Smith (2014) with minor
444 modifications (code available at https://bitbucket.org/ningwang83/Portulacineae). We describe the
445 procedure briefly below. An all-by-all BLASTN search was conducted using CDS data (-evalue 10 -
446 max_target_seqs 1000) and hits with a hit fraction > 0.49 (i.e. aligned length divided by query and by
447 subject length both > 0.4) were clustered using MCL 14-137 (-tf 'gq(5)' -I 1.4; van Dongen 2000). We
448 retrieved clusters including at least 26 species in Cactaceae, seven species in Montiaceae, seven species
449 in Didiereaceae, two species in each of Portulacaceae, Talinaceae and Molluginaceae, and one species
450 in each of Anacampserotaceae and outgroup as putative homolog groups (supplementary tables S1 and
451 S2, Supplementary Material online). Each cluster with less than 1000 sequences was aligned by
452 MAFFT v7 (–genafpair –maxiterate 1000; Katoh and Standley 2013) and aligned columns with more
453 than 90% missing data were removed using Phyutility v2.2.6 (-clean 0.1; Smith and Dunn 2008). A
454 maximum likelihood (ML) tree for each cluster was inferred in RAxML v8.1.22 (Stamatakis 2015)
455 with the GTR+GAMMA model. The six clusters containing more than 1000 sequences were aligned
456 with PASTA v1.6.4 (Mirarab et al. 2015), trimmed with "-clean 0.05" in Phyutility, and their ML trees
457 were inferred in FastTree v2.1.7 (Price et al. 2009, 2010) with the GTR model.

458 We trimmed the resulting homolog trees by 1) removing spurious terminal branches that are
459 longer than 0.5 (absolute cutoff), or longer than 0.4 and ten times greater than their sister taxon
460 (relative cutoff), 2) masking both monophyletic and paraphyletic tips of the same taxon to remove
461 possible isoforms, in-paralogs, or recent gene duplications, and 3) separating clades connected by
462 internal branches longer than 0.6; saving those with more than five taxa into different clusters as they
463 might represent alternate transcripts or in-paralogs. A new set of CDS sequence files were generated
464 based on the refining trees, and the process of tree building and filtering was repeated three times.

465 The PyPHLAWD package (https://github.com/FePhyFoFum/pyphlawd) was used to add data
466 from two additional species (*Anacampseros filamentosa* and *Basella alba*) during the third round of
467 homolog filtering process instead of conducting the time consuming all-by-all BLASTN. Specifically,
468 after modifying some default settings in the conf.py (i.e., "takeouttaxondups = False" to keep
469 duplication taxa; As to BLASTN hits filtering: length_limit = 0.6 to exclude sequences shorter than 60%
470 of the alignments, evalue_limit = 5 to keep hits with evalue less than 5, perc_identity = 20 to exclude
471 sequences with percentage of identity less than 20%.) and preparing the CDS files of the two species,
472 the add_internal_seqs_to_clusters.py was run twice (one species at a time) to BLAST target species'
473 CDS sequences against each homolog, and then added the hits that meet the above criteria to the
474 existing homologs and merge them into the existing alignments using MAFFT v7. In total, we obtained

475 8,592 final homolog clusters for inferring orthologs and conducting gene duplication analyses and GO
476 annotation.
477
478 **Orthology Inference and Species Tree Estimation**
479 We used the rooted tree (RT) method in Yang and Smith (2014) to extract orthologs from homolog
480 trees with *Beta vulgaris* as the outgroup. Sister clades with duplicate taxa were compared, and the side
481 with less taxa was removed. This procedure was carried out from root to tips iteratively on all
482 subclades until no duplicate taxa were present. Orthologous clades with more than 30 ingroup taxa
483 were retained and subject to calculation of taxon occupancy statistics (supplementary fig. S1,
484 Supplementary Material online). We further extracted orthologs with at least 76 taxa out of 82, aligned
485 and cleaned them with MAFFT (genafpair –maxiterate 1000) and Phyutility (-clean 0.3), and inferred
486 an ML tree for each ortholog using RAxML with GTR+GAMMA model. After removing taxa with
487 terminal branch length longer than 0.1 and 10 times greater than sister clade from 58 orthologs, 841
488 orthologs include at least 77 taxa and 500 aligned DNA characters were re-aligned with PRANK
489 v.140110 (Löytynoja and Goldman 2008) using default settings and trimmed in Phyutility (-clean 0.3).
490 The resulting alignments were used for phylogenetic tree reconstruction.
491 The species tree was inferred using two methods. First, a concatenated matrix was built with the
492 841 orthologous genes and a ML tree was estimated by using RAxML with the GTR+GAMMA model.
493 Node support was evaluated by 200 fast bootstrap replicates, partitioning by each gene. Second, we
494 constructed a ML tree and 200 fast bootstrap trees for each ortholog using RAxML with
495 GTR+GAMMA model. The MQSST tree was estimated in ASTRAL 4.10.12 (Mirarab et al. 2014)
496 using ML trees from RAxML, with uncertainty evaluated by 200 bootstrap replicates using a two-stage
497 multilocus bootstrap strategy (Seo 2008).
498
499 **Assessing conflicts among gene trees**
500 Conflicts among gene trees were assessed using the 841 rooted ortholog trees mapped onto the
501 concatenated ML tree (which has the same topology as the MQSST, see Results and Discussion).
502 When rooting orthologous trees with Phyx (Brown et al. 2017), the outgroup was chosen based on a
503 three-level preference (i.e., 1) using *Beta vulgaris* in priority, 2) using *Limeum aethiopicum* and/or
504 *Stegnosperma halimifolium* if *Beta* was not present in the tree, and 3) using *Sesuvium portulacastrum*,
505 *Delosperma echinatum*, *Anisomeria littoralis* and *Guapira obtusata* if the first three species were not
506 present). Phyparts (Smith SA et al. 2015) was used for comparing gene trees to the inferred species tree
507 topology, restricted to nodes with >= 70% bootstrap support in the gene trees. We explored conflict and

508 alternative topologies in more detail for three areas of the tree: 1) Cactaceae and relatives, 2) the early
509 diverging branch within Cactaceae, and 3) with the placement of Basellaceae and Didiereaceae. To do
510 this, we constructed constraint trees for each alternative topology, calculated likelihood scores for each
511 gene tree with the individual topology constraints, and then compared the likelihood scores for each
512 alternative resolution in each gene.
513
514 **Inference of gene and genome duplication**
515 To infer gene duplication, we first assessed clade support using the SH-like test (Anisimova et al. 2011)
516 in RAxML for each homolog tree. Nodes with SH-like support > 80 were considered in the gene
517 duplication inference. Using *Beta vulgaris* as outgroup, we extracted 8,332 rooted clusters (each
518 include 30 or more taxa, Yang et al. 2015), identified duplication events from them, and mapped those
519 events to the corresponding node or MRCA (if gene tree had missing taxa or a conflicting topology) on
520 the rooted concatenated ML tree using Phyparts. Gene duplication was recorded at a node if its two
521 children clades shared two or more taxa.
522 In addition to mapping phylogenetic locations of gene duplications, potential genome
523 duplication events were also inferred using the frequency distribution of synonymous substitution rate
524 (Ks plot) following the same process of Yang et al. (2015). Briefly, we reduced highly similar PEP
525 sequences (by CD-HIT: -c 0.99 -n 5) for each of the 82 species, and conducted an all-by-all BLASTP (-
526 evalue = 10, -max_target_seq = 20) and removed highly divergent hits with < 20% similarity (pident)
527 or < 50 aligned amino acids (nident). Sequences with ten or more hits were also removed to eliminate
528 large gene families. We then used the PEP and corresponding CDS from paralogous pairs to calculate
529 Ks values using pipeline
530 https://github.com/tanghaibao/biopipeline/tree/master/synonymous_calculation, which aligns
531 paralogous pairs with ClustalW (Larkin et al. 2007) and infers Ks values using codeml in PAML (Yang
532 2007) with Nei–Gojobori correction for multiple substitutions (Nei and Gojobori 1986). Peaks in the
533 distribution of Ks suggest the presence and relative timing of ancient genome duplication events
534 (Lynch and Conery 2000; Blanc and Wolfe 2004; Schlueter et al. 2004; Cannon et al. 2014).
535 To determine whether a potential WGD event occurred before or after a particular speciation
536 event, we calculated between-species Ks distribution using orthologous gene pairs (Cannon et al. 2014).
537 The procedure is similar as above, except that a reciprocal BLASTP was carried out between two
538 species instead of an all-by-all BLASTP within one taxon. Gene pairs with reciprocal best hits were
539 used for calculating the Ks distribution. The between-species Ks distribution was then compared to

540 within-species Ks in each species to determine the relative timing of the WGD versus the speciation
541 event.
542
543 **Annotation of expanded gene families**
544 To investigate gene function and GO (gene ontology) terms overrepresentation, the CDSs from
545 *Arabidopsis thaliana* (release-34) were obtained from EnsemblPlants and used as a BLASTN database,
546 while its gene IDs were used for GO analyses. We conducted two separate analyses. First, we identified
547 clades with large number of gene duplications [i.e., > 160 (2%) gene duplications as in node 15, 17 and
548 19 of Montiaceae, node 27 of Didieraceae, node 35, 39, 50 of Cactaceae, and node 71 of Portulacaceae,
549 fig. 1], conducted BLASTN (-evalue 10) of these duplication genes (ten randomly selected sequences
550 from each gene) in each clade against the database of *A. thaliana*. A random *A. thaliana* gene ID was
551 chosen when multiple query sequences from one gene returned divergent top hits, and the percentage
552 of identifiable genes were similar among clades (~35%, table 1). Assuming genes duplicated at the
553 origin of Portulacineae (Node 13) retained in all children clades, we use them as background to detect
554 potential overrepresentation of genes that recently duplicated in specific child clade using the GO
555 overrepresentation test of the online version of PANTHER (http://www.pantherdb.org). We corrected
556 for multiple tests using a Bonferroni correction. Here overrepresentation is unrelated to gene
557 expression level, instead, it refers to the overrepresentation of genes in a specific functional category in
558 one clade compared to another.
559 Second, we summarized each of the 8,592 homologous gene statistics such as the total number
560 of tips, the total number of taxa, clades with the maximum number of average taxon repeats, and the
561 taxon with the maximum number of tips. We then conducted manual annotation through BLASTX
562 against the nonredundant protein database in NCBI by using a relatively long and complete PEP
563 sequence from each of the top 20 genes with the highest total number of tips.
564
565 **Identification of lineage-specific selection on stress response genes**
566 To test if genes associated with abiotic stress response are under positive selection, we carried out
567 selection analyses on a targeted set of 29 genes that are previously known to be associated with cold
568 and/or drought adaptation, and also exist in our 8,592 homologous dataset (table 4). We first repeated
569 the homolog filtering processes as indicated above three more times to reduce potential assembly or
570 clustering errors of the chosen homologs. Their CDS sequences were then aligned based on the
571 corresponding peptide MAFFT alignments using phyx program (Brown et al. 2017). After trimming
572 the CDS alignment by Phyutility (-clean 0.1), a ML tree was built for each homolog in RAxML with

573 GTR+GAMMA model and rooted with outgroup species using the same three-level criteria (see above).
574 Lineage specific selection was calculated for each homologous gene in HyPhy v2.2.4 (Pond and Muse
575 2005) using an adaptive branch-site random effects likelihood test (aBSREL, Smith MD et al. 2015).
576 Generally, the ratio of the nonsynonymous to the synonymous substitutions ($\omega$ was used to measure the
577 extent pressure of natural selection, with $\omega > 1$ representing positive selection, $\omega \sim 1$ indicating neutral
578 evolution, and $\omega < 1$ representing purifying or negative selection (Yang 2006). aBSREL automatically
579 infers an appropriate model among branches. It allows non-synonymous substitution to vary across
580 sites, and synonymous rate can vary from branch to branch. This way $\omega$ varies among branches as well
581 as among sites. The aBSREL test first fits the MG94xREV nucleotide substitution model to estimate a
582 single $\omega$ for each branch, then greedily add $\omega$ categories to each branch after sorting them by length.
583 The optimal number of $\omega$ categories was selected according to AICc scores. In our analyses, all internal
584 branches were tested. The branches with episodic positive selection that show significant proportion of
585 sites with $\omega > 1$ were chosen with $p < 0.05$ after applying the Holm-Bonferroni multiple testing
586 correction. Finally, lineages subject to positive selection were labeled on each homologous tree. We
587 also summarized the number lineages under positive selection both on the whole tree and within each
588 focal clade (e.g., the major families within Portulacineae, fig. 1).
589
590 **Supplementary Material**
591 Supplementary data are available at Molecular Biology and Evolution online.
592
593 **Acknowledgments**
594 The authors thank Wynn Anderson, the Bureau of Land Management, the US Forest Service, and the
595 staff of Desert Botanical Garden, Sukkulenten-Sammlung Zürich, Cambridge University Botanic
596 Garden, Missouri Botanical Garden, and the Oberlin College Greenhouse for permission to collect
597 specimens, and thank Hilda Flores, Helga Ochoterena, and Norman Douglas for help with collecting.
598 The authors also thank Jeet Sukumaran for constructive discussion on gene tree discordance. Special
599 thanks to Oscar Vargas for a thorough review of the paper. This work was supported by NSF DEB
600 1354048 to SAS and NSF DEB 1352907 to MJM.
601
602 **References**
603 Abbasi F, Onodera H, Toki S, Tanaka H, Komatsu S. 2004. OsCDPK13, a calcium-dependent protein kinase
604 gene from rice, is induced by cold and gibberellin in rice leaf sheath. *Plant Mol Biol.* 55(4):541–552.

605 Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of
606 Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic
607 confidence. *Mol Biol Evol.* 20(2):255–266.
608 Anderson EF. 2001. The Cactus Family, Pentland, Oregon: Timber Press.
609 Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the
610 orders and families of flowering plants: APG IV. *Bot J Linean Soc.* 181(1):1–20.
611 Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods
612 demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.*
613 60(5):685–699.
614 Anton AM, Hernández-Hernández T, De-Nova A, Sosa V. 2014. Evaluating the phylogenetic position of the
615 monotypic family Halophytaceae (Portulacinae, Caryophyllales) based on plastid and nuclear molecular
616 data sets. *Bot Sci.* 92(3):351–361.
617 Applequist WL, Wagner WL, Zimmer EA, Nepokroeff M. 2006. Molecular evidence resolving the systematic
618 position of *Hectorella* (Portulacaceae). *Syst Bot.* 31(2):310–319.
619 Applequist WL, Wallace RS. 2000. Phylogeny of the Madagascan endemic family Didiereaceae. *Plant Syst Evol.*
620 221(3):157–166.
621 Applequist WL, Wallace RS. 2001. Phylogeny of the portulacaceous cohort based on ndhF sequence data. *Syst*
622 *Bot.* 26(2):406–419.
623 Applequist WL, Wallace RS. 2003. Expanded circumscription of Didiereaceae and its division into three
624 subfamilies. *Adansonia* 25(1):13–16.
625 Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011.
626 Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad*
627 *Sci USA.* 108(20):8379–8384.
628 Bárcenas RT, Yesson C, Hawkins JA. 2011 Molecular systematics of the Cactaceae. *Cladistics* 27:470–489
629 Barthlott W, Hunt DR. 1993. Cactaceae. In Flowering Plants Dicotyledons Springer Berlin Heidelberg. p. 161–
630 197.
631 Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of
632 duplicate genes. *Plant Cell* 16(7):1667–1678.
633 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
634 *Bioinformatics* 30(15):2114–2120.
635 Borland AM, Griffiths H, Hartwell J, Smith JA. 2009. Exploiting the potential of plants with crassulacean acid
636 metabolism for bioenergy production on marginal lands. *J Exp Bot.* 60(10):2879–2896.
637 Brandt B, Brodsky DE, Xue S, Negi J, Iba K, Kangasjärvi J, Ghassemian M, Stephan AB, Hu H, Schroeder JI.
638 2012. Reconstitution of abscisic acid activation of SLAC1 anion channel by CPK6 and OST1 kinases
639 and branched ABI1 PP2C phosphatase action. *Proc Natl Acad Sci USA.* 109(26):10593–10598.

640 Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R. 2004. Polyploidy in arctic
641 plants. *Bio J Linnean Soc.* 82:521–536.
642 Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GK, Moore MJ, Smith
643 SA. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in
644 Caryophyllales. *New Phytol.* 207:1170–1180.
645 Brown JM, Thomson RC. 2017. Bayes factors unmask highly variable information content, bias, and extreme
646 influence in phylogenomic analyses. *Syst Biol.* 66:517–530.
647 Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33(12):1886–1888.
648 Bruyns PV, Oliveira-Neto M, Melo-de-Pinna GF, Klak C. 2014. Phylogenetic relationships in the Didiereaceae
649 with special reference to subfamily Portulacarioideae. *Taxon* 63(5):1053–1064.
650 Butterworth CA, Cota-Sanchez JH, Wallace RS. 2002. Molecular systematics of tribe Cacteae (Cactaceae:
651 Cactoideae): a phylogeny based on rpl16 intron sequence variation. *Syst Bot.* 27(2):257–270.
652 Butterworth CA, Wallace RS. 2005. Molecular phylogenetics of the leafy cactus genus *Pereskia* (Cactaceae).
653 *Syst Bot.* 30(4):800–808.
654 Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart Jr CN, Rolf
655 M, Kutchan T. 2014. Multiple polyploidy events in the early radiation of nodulating and nonnodulating
656 legumes. *Mol Biol Evol.* 32(1):193–210.
657 Chan KX, Wirtz M, Phua SY, Estavillo GM, Pogson BJ. 2013. Balancing metabolites in drought: the sulfur
658 assimilation conundrum. *Trends Plant Sci.* 18(1):18–29.
659 Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought–from genes to the whole
660 plant. *Funct Plant Biol.* 30(3):239–264.
661 Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK. 2003. ICE1: a regulator of cold-
662 induced transcriptome and freezing tolerance in Arabidopsis. *Genes Dev.* 17(8):1043–1054.
663 Christin PA, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK,
664 Hancock L, Edwards EJ. 2014. Shared origins of a key enzyme during the evolution of C4 and CAM
665 metabolism. *J Exp Bot.* 65:3609–3621.
666 Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 Photosynthesis Evolved in Grasses via
667 Parallel Adaptive Genetic Changes. *Curr Biol.* 17:1241–1247.
668 Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K,
669 Barakat A, Albert VA. 2006. Widespread genome duplications throughout the history of flowering
670 plants. *Genome Res.* 16(6):738–749.
671 Cushman JC. 2001. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments.
672 *Plant Physiol.* 127:1439–1448.
673 Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent.
674 *Trends Ecol Evol.* 24(6):332–340.

675 Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR,
676 Stracke R, Reinhardt R, Goesmann A. 2014. The genome of the recently domesticated crop plant sugar
677 beet (*Beta vulgaris*). *Nature* 505(7484):546–549.
678 Edwards EJ, Nyffeler R, Donoghue MJ. 2005. Basal cactus phylogeny: implications of *Pereskia* (Cactaceae)
679 paraphyly for the transition to the cactus life form. *Am J Bot.* 92(7):1177–1188.
680 Edwards EJ, Ogburn RM. 2012. Angiosperm responses to a low-CO2 world: CAM and C4 photosynthesis as
681 parallel evolutionary trajectories. *Int J of Plant Sci.* 173(6):724–733.
682 Edwards G, Walker D. 1983. C3, C4: mechanisms, and cellular and environmental regulation, of photosynthesis.
683 University of California Press.
684 Emms DM, Covshoff S, Hibberd JM, Kelly S. 2016. Independent and parallel evolution of new genes by gene
685 duplication in two origins of C4 photosynthesis provides new insight into the mechanism of phloem
686 loading in C4 species. *Mol Biol Evol.* 33(7):1796–1806.
687 Ferreyra MLF, Andreo CS, Podestá FE. 2003. Purification and physical and kinetic characterization of a
688 photosynthetic NADP-dependent malic enzyme from the CAM plant Aptenia cordifolia. *Plant Sci.*
689 164(1):95–102.
690 Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation
691 in Arabidopsis thaliana. *Science* 334(6052):86–89.
692 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data.
693 *Bioinformatics* 28(23):3150–3152.
694 Geiger D, Scherzer S, Mumm P, Marten I, Ache P, Matschi S, Liese A, Wellmann C, Al-Rasheid KA, Grill E,
695 Romeis T, Hedrich R. 2010. Guard cell anion channel SLAC1 is regulated by CDPK protein kinases
696 with distinct Ca2+ affinities. *Proc Natl Acad Sci USA.* 107(17):8023–8028.
697 Georges F, Das S, Ray H, Bock C, Nokhrina K, Kolla VA, Keller W. 2009. Over-expression of *Brassica napus*
698 phosphatidylinositol-phospholipase C2 in canola induces significant changes in gene expression and
699 phytohormone distribution patterns, enhances drought tolerance and promotes early flowering and
700 maturation. *Plant Cell Environ.* 32(12):1664–1681.
701 Gibson AC, Nobel PS. 1990. The cactus primer. Harvard University Press
702 Gibson AC. 1996. Structure–function relations of warm desert plants. New York, NY, USA: Springer.
703 Golldack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: unraveling the
704 signaling networks. *Front Plant Sci.* 5:151.
705 Gosti F, Beaudoin N, Serizet C, Webb AA, Vartanian N, Giraudat J. 1999. ABI1 protein phosphatase 2C is a
706 negative regulator of abscisic acid signaling. *Plant Cell* 11(10):1897–1909.
707 Gowik U, Bräutigam A, Weber KL, Weber APM, Westhof P. 2011. Evolution of C4 photosynthesis in the genus
708 *Flaveria*: how many and which genes does it take to make C4? *Plant Cell* 23:2087–2105.

709 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R,
710 Zeng Q, Chen Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference
711 genome. *Nat Biotechnol.* 29(7):644–652.
712 Griffith MP, Porter JM. 2009. Phylogeny of Opuntioideae (Cactaceae). *International J Plant Sci.* 170:107–116.
713 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M,
714 MacManes MD. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity
715 platform for reference generation and analysis. *Nature Protoc.* 8(8):1494–1512.
716 Hernández-Hernández T, Hernández HM, De-Nova JA, Puente R, Eguiarte LE, Magallón S. 2011. Phylogenetic
717 relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot.*
718 98(1):44–61.
719 Hernández-Ledesma P, Berendsohn WG, Borsch T, Mering SV, Akhani H, Arias S, Castañeda-Noa I, Eggli U,
720 Eriksson R, Flores-Olvera H, Fuentes-Bazán S. 2015. A taxonomic backbone for the global synthesis of
721 species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45(3):281–383.
722 Hernández-Hernández T, Brown JW, Schlumpberger BO, Eguiarte LE, Magallón S. 2014. Beyond aridification:
723 multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New*
724 *Phytol.* 202(4):1382–1397.
725 Hershkovitz MA, Zimmer EA. 1997. On the evolutionary origins of the cacti. *Taxon* 46(2):217–232.
726 Hochstrasser M. 2009. Origin and function of ubiquitin-like proteins. *Nature* 458(7237):422–429.
727 Huang Q, Wang Y, Li B, Chang J, Chen M, Li K, Yang G, He G. 2015. TaNAC29, a NAC transcription factor
728 from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* 15(1):268.
729 Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI. 2010. Early abscisic acid signal transduction
730 mechanisms: newly discovered components and newly emerging questions. *Genes Dev.* 24(16):1695–
731 1708.
732 Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh
733 A. 2014. Whole-genome analyses resolve early branches in the tree of life of modern
734 birds. *Science* 346(6215):1320–1331.
735 Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y, Kim M, Reuzeau C, Kim JK. 2010. Root-specific
736 expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought
737 conditions. *Plant Physiol.* 153:185–197
738 Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H,
739 Soltis PS, Soltis DE. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–
740 100.
741 Jiang Y, Deyholos MK. 2009. Functional characterization of Arabidopsis NaCl-inducible WRKY25 and
742 WRKY33 transcription factors in abiotic stresses. *Plant Mol Biol.* 69(1–2):91–105.
743 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
744 performance and usability. *Mol Biol Evol.* 30(4):772–780.

745 Kiang JG, Tsokos GC. 1998. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology.
746 *Pharmacol Ther.* 80(2):183–201.
747 Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in
748 response to salt, osmotic, and cold stress. *Plant Physiol.* 130(4):2129–2141.
749 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM,
750 Wilm A, Lopez R, Thompson JD. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*
751 23(21):2947–2948.
752 Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson
753 DW, McCombie WR, Martienssen RA. 2011. A functional phylogenomic view of the seed plants. *PLoS*
754 *Genet.* 7(12):e1002411.
755 Li H, Chang J, Zheng J, Dong Y, Liu Q, Yang X, Wei C, Zhang Y, Ma J, Zhang X. 2017. Local melatonin
756 application induces cold tolerance in distant organs of *Citrullus lanatus* L. via long distance
757 transport. *Sci Rep.* 7:40858.
758 Li WH. 1997. Molecular evolution. Sinauer Associates, Sunderland, MA
759 Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA
760 sequence data. *Evolution* 62(8):2080–2091.
761 Liu L, Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a
762 species phylogeny using estimated gene tree distributions. *Syst Biol.* 56(3):504–514.
763 Liu Z, Jia Y, Ding Y, Shi Y, Li Z, Guo Y, Gong Z, Yang S. 2017. Plasma Membrane CRPK1-Mediated
764 Phosphorylation of 14-3-3 Proteins Induces Their Nuclear Import to Fine-Tune CBF Signaling during
765 Cold Response. *Mol Cell* 66(1):117–128.
766 Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and
767 evolutionary analysis. *Science* 320(5883):1632–1635.
768 Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*
769 290(5494):1151–1155.
770 Maddison WP. 1997. Gene trees in species trees. *Syst Biol.* 46:523–536
771 Manavella PA, Arce AL, Dezar CA, Bitton F, Renou JP, Crespi M, Chan RL. 2006. Cross-talk between ethylene
772 and drought signaling pathways is mediated by the sunflower Hahb-4 transcription factor. *Plant J.*
773 48(1):125–137.
774 Masoomi-Aladizgeh F, Aalami A, Esfahani M, Aghaei MJ, Mozaffari K. 2015. Identification of CBF14 and
775 NAC2 genes in *Aegilops tauschii* associated with resistance to freezing stress. *Appl Biochem Biotechnol.*
776 176(4):1059–1070.
777 Mauseth JD. 2006. Structure–function relationships in highly modified shoots of Cactaceae. *Ann Bot.* 98(5):901–
778 926.

779 Merlot S, Gosti F, Guerrier D, Vavasseur A, Giraudat J. 2001. The ABI1 and ABI2 protein phosphatases 2C act
780 in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* 25(3):295–303.
781 Mirarab S, Nguyen N, Guo S, Wang LS, Kim J, Warnow T. 2015. PASTA: ultra-large multiple sequence
782 alignment for nucleotide and amino-acid sequences. *J Comput Biol.* 22(5):377–386.
783 Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale
784 coalescent-based species tree estimation. *Bioinformatics* 30(17):i541–i548.
785 Mishra G, Zhang W, Deng F, Zhao J, Wang X. 2006. A bifurcating pathway directs abscisic acid effects on
786 stomatal closure and opening in Arabidopsis. *Science* 312(5771):264–266.
787 Moore A, de Vos JM, Hancock LP, Goolsby E, Edwards EJ. 2017. Targeted Enrichment of Large Gene Families
788 for Phylogenetic Inference: Phylogeny and Molecular Evolution of Photosynthesis Genes in the
789 Portullugo (Caryophyllales). *Syst Biol.* doi:10.1093/sysbio/syx078.
790 Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. NAC transcription factors in
791 plant abiotic stress responses. *Biochim Biophys Acta* 1819(2):97–103.
792 Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 2014. The transcriptional regulatory network in the
793 drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front*
794 *Plant Sci.* 5:170.
795 Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous
796 nucleotide substitutions. *Mol Biol Evol.* 3(5):418–426.
797 Nyffeler R, Eggli U. 2010a. Disintegrating Portulacaceae: a new familial classification of the suborder
798 Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* 59(1):227–240.
799 Nyffeler R, Eggli U. 2010b. A farewell to dated ideas and concepts –molecular phylogenetics and a revised
800 suprageneric classification of the family Cactaceae. *Schumannia* 6:109–149.
801 Nyffeler R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from
802 trnK/matK and trnL-trnF sequences. *Am J Bot.* 89:312–326.
803 Nyffeler R. 2007. The closest relatives of cacti: insights from phylogenetic analyses of chloroplast and
804 mitochondrial sequences with special emphasis on relationships in the tribe Anacampseroteae. *Am J Bot.*
805 94(1):89–101.
806 Ocampo G, Columbus JT. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales), including
807 insights into photosynthetic diversification and historical biogeography. *Am J Bot.* 97(11):1827–1847.
808 Ogburn RM, Edwards EJ. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm
809 clade Montiaceae. *Mol Phylogenet Evol.* 92:181–192.
810 Oh JE, Kwon Y, Kim JH, Noh H, Hong SW, Lee H. 2011. A dual role for MYB60 in stomatal regulation and
811 root growth of *Arabidopsis thaliana* under drought stress. *Plant Mol Biol.* 77(1-2):91–103.
812 Pan Y, Seymour GB, Lu C, Hu Z, Chen X, Chen G. 2012. An ethylene response factor (ERF5) promoting
813 adaptation to drought and salt tolerance in tomato. *Plant Cell Rep.* 31(2):349–360.

814 Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification.
815 *Hydrol Earth Syst Sci.* 11:1633–1644.
816 Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol*
817 *Evol.* 21:1455–1458.
818 Pond, S.L.K. and Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. In: Statistical methods in
819 molecular evolution. Springer, New York, NY. p. 125–181.
820 Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead
821 of a distance matrix. *Mol Biol Evol.* 26(7):1641–1650.
822 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large
823 alignments. *PloS One* 5(3):e9490.
824 Qin F, Sakuma Y, Li J, Liu Q, Li YQ, Shinozaki K, Yamaguchi-Shinozaki K. 2004. Cloning and functional
825 analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea*
826 *mays* L. *Plant Cell Physiol.* 45(8):1042–1052.
827 Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han KL, Harshman J, Huddleston CJ,
828 Kingston S, et al. 2017. Why Do Phylogenomic Data Sets Yield Conflicting Trees? Data Type
829 Influences the Avian Tree of Life more than Taxon Sampling. *Syst Biol.* 66:857–879.
830 Sajeevan RS, Nataraja KN, Shivashankara KS, Pallavi N, Gurumurthy DS, Shivanna MB. 2017. Expression of
831 Arabidopsis SHN1 in Indian mulberry (*Morus indica* L.) increases leaf surface wax content and reduces
832 post-harvest water loss. *Front Plant Sci.* 8:418.
833 Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC. 2004. Mining EST databases to
834 resolve evolutionary events in major crop species. *Genome* 47(5):868–876.
835 Seo TK. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol Biol Evol.*
836 25(5):960-971.
837 Shen X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a
838 handful of genes. *Nature Ecol Evol.* 1:1–10.
839 Silvera K, Winter K, Rodriguez BL, Albion RL, Cushman JC. 2014. Multiple isoforms of phosphoenol-pyruvate
840 carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid
841 metabolism. *J Exp Bot.* 65(13):3623–3636.
842 Smith JAC, Winter K. 1996. Taxonomic distribution of crassulacean acid metabolism. In: Winter K, Smith JAC,
843 eds. Crassulacean acid metabolism: biochemistry, ecophysiology and evolution. Berlin/Heidelberg:
844 Springer-Verlag. p. 427–436.
845 Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an
846 adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol*
847 *Biol Evol.* 32(5):1342–1353.

848 Smith SA, Brown JW, Yang Y, Bruenn R, Drummond CP, Brockington SF, Walker JF, Last N, Douglas NA,
849 Moore MJ. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.* 217(2):836–
850 854.
851 Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data.
852 *Bioinformatics* 24(5):715–716.
853 Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict,
854 concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* 15(1):150.
855 Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB,
856 Moore MJ, Carlsward BS, Bell CD. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot.*
857 98(4):704–730.
858 Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588.
859 Soltis PS, Soltis DS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl*
860 *Acad Sci USA*. 97:7051–7057.
861 Song L, Florea L. 2015. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads.
862 *GigaScience* 4:48
863 Song H, Zhao R, Fan P, Wang X, Chen X, Li Y. 2009. Overexpression of AtHsp90. 2, AtHsp90. 5 and AtHsp90.
864 7 in Arabidopsis thaliana enhances plant sensitivity to salt and drought stresses. *Planta* 229(4):955–964.
865 Stamatakis A. 2015. Using RAxML to infer phylogenies. *Curr Protoc Bioinformatics.* 51:6.14.1-14.
866 Stebbins GL. 1971. Chromosomal evolution in higher plants. Addison Wesley, London. p. 216.
867 Stoughton TR, Jolles DD, O'Quinn RL. 2017. The western spring beauties, *Claytonia lanceolata* (Montiaceae): a
868 review and revised taxonomy for California. *Syst Bot.* 42(2):283–300.
869 van den Bergh E, Külahoglu C, Bräutigam A, Hibberd JM, Weber AP, Zhu XG, Schranz ME. 2014. Gene and
870 genome duplications and the origin of C4 photosynthesis: Birth of a trait in the Cleomaceae. *Curr Plant*
871 *Biol.* 1:2–9.
872 van Dongen S. 2000. Graph clustering by flow simulation [PhD thesis]. [Utrecht, The Netherlands]: University
873 of Utrecht.
874 Vandeleur RK, Mayo G, Shelden MC, Gilliham M, Kaiser BN, Tyerman SD. 2009. The role of plasma
875 membrane intrinsic protein aquaporins in water transport through roots: diurnal and drought stress
876 responses reveal different strategies between isohydric and anisohydric cultivars of grapevine. *Plant*
877 *Physiol.* 149(1):445–460.
878 Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions revisited.
879 *Mol Biol Evol.* 30(1):177–90.
880 Vázquez-Sánchez M, Terrazas T, Arias S, Ochoterena H. 2013. Molecular phylogeny, origin and taxonomic
881 implications of the tribe Cacteae (Cactaceae). *Sys Biodivers.* 11(1):103-116.

882 Vivek PJ, Tuteja N, Soniya EV. 2013. CDPK1 from ginger promotes salinity and drought stress tolerance
883 without yield penalty by improving growth and photosynthesis in *Nicotiana tabacum*. *PLoS*
884 *One* 8(10):e76392.
885 Walker JF, Yang Y, Feng T, Timoneda A, Mikenas J, Hutchinson V, Edwards C, Wang N, Ahluwalia S, Olivieri
886 J, et al. 2018a. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical
887 homology inference provide further insight to the evolution of Caryophyllales. Am J Bot. special issue
888 on the Plant Tree of Life, in press.
889 Walker JF, Brown JW, Smith, SA. 2018b. Analyzing contentious relationships and outlier genes in
890 phylogenomics. BioRixv doi: https://doi.org/10.1101/115774.
891 Walker JF, Yang Y, Moore MJ, Mikenas J, Timoneda A, Brockington SF, Smith SA. 2017. Widespread
892 paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales.
893 *Am J Bot.* 104(6):858–867.
894 Wallace RS, Dickie SL. 2002. Systematic implications of chloroplast DNA sequence variation in subfam.
895 Opuntioideae (Cactaceae). *Succ. Pl. Res.* 6: 9–24.
896 Wang F, Chen HW, Li QT, Wei W, Li W, Zhang WK, Ma B, Bi YD, Lai YC, Liu XL, Man WQ. 2015.
897 GmWRKY27 interacts with GmMYB174 to reduce expression of GmNAC29 for stress tolerance in
898 soybean plants. *Plant J.* 83(2):224–236.
899 Wang L, Jin X, Li Q, Wang X, Li Z, Wu X. 2016. Comparative Proteomics Reveals that Phosphorylation of β
900 Carbonic Anhydrase 1 Might be Important for Adaptation to Drought Stress in *Brassica napus*. *Sci Rep.*
901 6:39024.
902 Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones
903 in the abiotic stress response. *Trends Plant Sci.* 9(5):244–252.
904 Watson A. 1979. Gypsum crusts in deserts. *J Arid Environ*. 2(1):3–21.
905 Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *Am J Bot.* 102:1753–1756.
906 Williams BP, Aubry S, Hibberd JM. 2012. Molecular evolution of genes recruited into C4
907 photosynthesis. *Trends Plant Sci.* 17(4):213–220.
908 Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of
909 polyploid speciation in vascular plants. *Proc Natl Acad Sci USA.* 106:13875–13879.
910 Xie Z, Li D, Wang L, Sack FD, Grotewold E. 2010. Role of the stomatal development regulators FLP/MYB88
911 in abiotic stress responses. *Plant J.* 64(5):731–739.
912 Yang L, Ji W, Zhu Y, Gao P, Li Y, Cai H, Bai X, Guo D. 2010. GsCBRLK, a calcium/calmodulin-binding
913 receptor-like kinase, is a positive regulator of plant tolerance to salt and ABA stress. *J Exp Bot.*
914 61(9):2519–2533.
915 Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, Smith SA. 2018. Improved transcriptome
916 sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two
917 allopolyploidy events. *New Phytol.* 217(2):855–870

918 Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y,
919 Sage RF. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using
920 transcriptome sequencing. *Mol Biol Evol.* 32(8):2001–2014.
921 Yang Y, Moore MJ, Brockington SF, Timoneda A, Feng T, Marx HE, Walker JF, Smith SA. 2017. An efficient
922 field and laboratory workflow for plant phylotranscriptomic projects. *Appl Plant Sci.* 5(3):1600128
923 Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage
924 genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol Biol Evol.* 31(11):3081–
925 3092.
926 Yang Z. 2006. Computational Molecular Evolution. Oxford: Oxford University Press. p. 357
927 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
928 Yang T, Shad Ali G, Yang L, Du L, Reddy AS, Poovaiah BW. 2010. Calcium/calmodulin-regulated receptor-
929 like kinase CRLK1 interacts with MEKK1 in plants. *Plant Signal Behav*. 5(8):991–994.
930 Yoshida R, Umezawa T, Mizoguchi T, Takahashi S, Takahashi F, Shinozaki K. 2006. The regulatory domain of
931 SRK2E/OST1/SnRK2. 6 interacts with ABI1 and integrates abscisic acid (ABA) and osmotic stress
932 signals controlling stomatal closure in Arabidopsis. *J Biol Chem.* 281(8):5310–5318.
933 Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K.
934 2010. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-
935 dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant*
936 *J.* 61(4):672–685.
937 Yuan LL, Zhang M, Yan X, Bian YW, Zhen SM, Yan YM. 2016. Dynamic phosphoproteome analysis of
938 seedling leaves in *Brachypodium distachyon* L. reveals central phosphorylated proteins involved in the
939 drought stress response. *Sci Rep.* 6:35280.
940 Zhu JK. 2016. Abiotic stress signaling and responses in plants. *Cell* 167(2):313–324.
941 Zou JJ, Wei FJ, Wang C, Wu JJ, Ratnasekera D, Liu WX, Wu WH. 2010. Arabidopsis calcium-dependent
942 protein kinase CPK10 functions in abscisic acid-and Ca2+-mediated stomatal regulation in response to
943 drought stress. *Plant Physiol.* 154(3):1232–1243.
944
945
946

947 Table 1. Summary of clade-specific duplication genes and identifiable GO terms

948

| Node number | Duplicated homologs | Identifiable AtIDs[1] | Mapped IDs[2] | Percent identifiability | Duplication events |
|---|---|---|---|---|---|
| Node 13 | 761 | 341 | 329 | 0.45 | 1118 |
| Node 15 | 167 | 94 | 90 | 0.56 | 187 |
| Node 17 | 1516 | 459 | 457 | 0.30 | 1589 |
| Node 19 | 376 | 117 | 117 | 0.31 | 388 |
| Node 26 | 111 | 61 | 54 | 0.55 | 111 |
| Node 27 | 3383 | 1082 | 1065 | 0.32 | 4171 |
| Node 35 | 534 | 217 | 212 | 0.41 | 612 |
| Node 39 | 474 | 161 | 159 | 0.34 | 532 |
| Node 50 | 276 | 97 | 95 | 0.35 | 315 |
| Node 71 | 155 | 55 | 55 | 0.35 | 170 |

949 NOTE.—Node number corresponds to Fig. 1. [1]Identifiable *Arabidopsis* IDs;

950 [2]Genes that can be mapped in PANTHER.

951 Table 2. GO overrepresentation analyses on specific clades

952

| Node number | 13 | 15 | | | | 27 | | | | 35 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene Ontology term | Ob. | Ob. | Exp. | +/- | p | Ob. | Exp. | +/- | p | Ob. | Exp. | +/- | p |
| chromatin binding (GO:0003682) | 1 | 0 | 0.28 | - | 1 | *11* | *3.17* | + | *0.028* | 1 | 0.64 | + | 1 |
| translation regulator activity (GO:0045182) | 1 | 0 | 0.28 | - | 1 | **12** | **3.17** | **+** | **0.007** | 3 | 0.64 | + | 1 |
| sequence-specific DNA binding TF[1] activity (GO:0003700) | 4 | 1 | 1.1 | - | 1 | **29** | **12.69** | **+** | **0.003** | 5 | 2.55 | + | 1 |
| calcium ion binding (GO:0005509) | 1 | **5** | **0.28** | **+** | **6E-4** | **15** | **3.17** | **+** | **8E-5** | 3 | 0.64 | + | 1 |
| hydrolase activity, acting on ester bonds (GO:0016788) | 7 | 4 | 1.93 | + | 1 | 40 | 22.21 | + | 0.056 | *14* | *4.45* | + | *0.028* |
| rRNA metabolic process (GO:0016072) | 1 | 0 | 0.28 | - | 1 | *11* | *3.17* | + | *0.041* | 2 | 0.64 | + | 1 |
| sulfur compound metabolic process (GO:0006790) | 1 | 0 | 0.28 | - | 1 | *11* | *3.17* | + | *0.041* | 1 | 0.64 | + | 1 |
| regulation of transcription from RNA P-II[2] promoter (GO:0006357) | 2 | 0 | 0.55 | - | 1 | **18** | **6.35** | **+** | **0.01** | 0 | 1.27 | - | 1 |
| transcription, DNA-dependent (GO:0006351) | 6 | 2 | 1.65 | + | 1 | **39** | **19.04** | **+** | **0.003** | 3 | 3.82 | - | 1 |

953 NOTE.—Node number corresponds to Fig. 1. Ob. : Observed number; Exp. : Expected number; +/-: over or under representative. Italic: $p < 0.05$, Bold: $p < 0.01$. [1]TF: transcription factor; [2]RNA P-II: RNA polymerase II.

954

955 Table 3. Portulacineae clades with the highest total number of tips

956

| clusterID | Ntip | Ntaxa | Represented ID | Annotation |
|---|---|---|---|---|
| cluster1_2rr_1rr_2rr | 949 | 82 | AT5G62690.1 | tubulin beta chain [*Spinacia oleracea*] |
| cluster2_1rr_1rr_1rr | 871 | 82 | AT5G02500.2 | heat shock cognate 70 kDa protein-like [*Nicotiana tabacum*] |
| cluster4_1rr_1rr_1rr | 698 | 82 | AT5G09810.1 | Actin-11, NBD sugar-kinase HSP70 actin [*Chenopodium quinoa*] |
| cluster3_3rr_1rr_1rr | 670 | 81 | AT1G55060.1 | Polyubiquitin-A [*Triticum urartu*] |
| cluster5_1rr_1rr_1rr | 655 | 82 | AT2G18960.3 | plasma membrane H+ ATPase 9 [*Sesuvium portulacastrum*] |
| cluster8_1rr_1rr_1rr | 582 | 82 | AT5G19780.1 | tubulin alpha-3 chain [*Vitis vinifera*] |
| cluster12_1rr_1rr_1rr | 565 | 82 | AT3G14940.2 | phosphoenolpyruvate carboxylase [*Suaeda aralocaspica*] |
| cluster43_1rr_1rr_1rr | 488 | 81 | AT3G08690.1 | ubiquitin-conjugating enzyme E2 28 [*Solanum lycopersicum*] |
| cluster18_1rr_1rr_1rr | 472 | 82 | AT4G14350.4 | serine/threonine-protein kinase tricorner isoform X1 [*Beta vulgaris vulgaris*] |
| cluster22_1rr_1rr_1rr | 467 | 82 | AT2G19900.1 | NADP-dependent malic enzyme [*Hylocereus undatus*] |
| cluster15_1rr_1rr_1rr | 463 | 82 | AT5G37310.2 | transmembrane 9 superfamily member 3 [*Chenopodium quinoa*] |
| cluster17_1rr_1rr_1rr | 462 | 82 | AT2G21050.1 | auxin transporter-like protein 2 [*Chenopodium quinoa*] |
| cluster19_1rr_1rr_1rr | 461 | 82 | AT2G37170.2 | plasma membrane intrinsic protein 2 [*Sesuvium portulacastrum*] |
| cluster7_1rr_1rr_1rr | 456 | 81 | AT1G26190.2 | uridine-cytidine kinase C isoform X1 [*Beta vulgaris vulgaris*] |
| cluster26_1rr_1rr_1rr | 435 | 82 | AT4G01850.1 | S-adenosylmethionine synthase 1 [*Beta vulgaris vulgaris*] |
| cluster14_1rr_1rr_1rr | 467 | 82 | AT4G30270.2 | probable xyloglucan endotransglucosylase/hydrolase protein 23 [*Beta vulgaris vulgaris*] |
| cluster6_2rr_1rr_1rr | 530 | 81 | AT4G32880.1 | homeobox-leucine zipper protein ATHB-15 isoform X1 [*Beta vulgaris vulgaris*] |
| cluster9_1rr_1rr_1rr | 527 | 81 | AT3G29030.2 | expansin-A10-like [*Chenopodium quinoa*] |
| cluster16_1rr_1rr_1rr | 438 | 82 | AT5G06600.3 | ubiquitin carboxyl-terminal hydrolase 12 [*Beta vulgaris vulgaris*] |
| cluster37_1rr_1rr_1rr | 435 | 82 | AT1G22300.2 | 14-3-3 protein 10-like [*Chenopodium quinoa*] |

957 NOTE.—Ntip: number of tips, Ntaxa: number of taxa

958 Table 4. Selection analyses on genes associated with cold and drought

959

| clusterID | Gene | TN | TS | M7 | Cac35 | Did27 | Mont | Port | SH80 | No support constraint | References |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cluster1007 | CDPK8 | 184 | 2 | 0 | 1 | 0 | 0 | 0 | 13,27,71 | 2,13,27,35,71 | Wang et al. 2016 |
| cluster1037 | WRKY33 | 165 | 13 | 1 | 4 | 2 | 3 | 1 | 17,27 | 0,17,27 | Jiang and Deyholos 2009 |
| cluster112 | HSP70 | 270 | 10 | 0 | 3 | 1 | 1 | 2 | 0,19,27,33,49,51,52 | 0,19,27,33,39,49,51,50,52,56 | Kiang and Tsokos 1998 |
| cluster1131 | ICE1 | 194 | 10 | 0 | 4 | 1 | 3 | 0 | 17,20,27,38 | 2,17,20,27,38,51,50 | Chinnusamy et al. 2003 |
| **cluster1181** | **ABI1/PP2Cfamily** | 151 | 17 | 1 | 6 | 2 | 4 | 0 | 27 | 0,15,27,40 | Merlot et al. 2001 |
| cluster1248 | NAC10 | 158 | 16 | 1 | 6 | 4 | 3 | 1 | 27,39 | 2,19,27,39 | Jeong et al. 2010 |
| cluster1349 | SNRK2.5 | 156 | 5 | 0 | 2 | 1 | 0 | 1 | 39 | 0,27,34,35,39 | Zhu 2016 |
| cluster1418 | NAC29 | 163 | 13 | 0 | 7 | 2 | 2 | 0 | 13 | 13,17,27,39,44,50 | Wang et al. 2015 |
| cluster1864 | CDPK13 | 156 | 3 | 0 | 0 | 0 | 0 | 2 | 27 | 2,27,39 | Abbasi et al. 2004 |
| cluster286 | MYB124 (FLP) | 178 | 5 | 0 | 2 | 1 | 0 | 1 | 39 | 6,17,19,27,39 | Xie et al. 2010 |
| cluster3368 | SNRK2.4 | 125 | 1 | 0 | 0 | 0 | 0 | 0 | 17,27 | 2,15,17,27,30 | Zhu 2016 |
| cluster3966 | CDPK26 | 117 | 3 | 0 | 1 | 1 | 0 | 1 | 0,17,27 | 0,17,27,39 | Yuan et al. 2016 |
| cluster4535 | PLC2 | 108 | 3 | 0 | 1 | 0 | 1 | 0 | 17,19,35 | 0,17,19,27,35 | Georges et al. 2009 |
| cluster5012 | CDPK10 | 85 | 1 | 0 | 1 | 0 | 0 | 0 | 27 | 0,27,60 | Zou et al. 2010 |
| cluster5640 | AREB1 | 92 | 4 | 0 | 1 | 1 | 1 | 0 | 17,27 | 17,27 | Yoshida et al. 2010 |
| cluster579 | NAC2 | 157 | 4 | 0 | 2 | 0 | 0 | 1 | 27,39 | 0,27,39 | Masoomi-Aladizgeh et al. 2015 |
| cluster582 | CDPK18 | 196 | 13 | 1 | 6 | 1 | 0 | 0 | 2,13,21,35,57 | 2,13,18,21,24,27,35,40,47,57 | Li et al. 2017 |
| cluster584 | CRLK1 | 170 | 2 | 0 | 2 | 0 | 0 | 0 | 27 | 13,27,39,50 | Yang et al. 2010 |
| cluster59 | HSP83 | 305 | 5 | 0 | 1 | 1 | 0 | 0 | 0,13,17,27,34,76 | 0,13,15,17,19,24,27,34,39,50,76 | Wang et al. 2004 |
| cluster6422 | MYB60 | 87 | 4 | 0 | 2 | 0 | 0 | 1 | 27 | 27,40 | Oh et al. 2011 |
| cluster643 | HSP90-5 | 213 | 8 | 1 | 2 | 2 | 1 | 1 | 26,27,34 | 2,19,26,27,34,39,71 | Song et al. 2009 |
| cluster6979 | CDPK1 | 85 | 7 | 0 | 4 | 2 | 1 | 0 | --- | 27 | Vivek et al. 2013 |
| cluster708 | SNRK2.6 | 180 | 2 | 0 | 2 | 0 | 0 | 0 | 27,39 | 0,27,39,50 | Yoshida et al. 2006 |
| cluster7152 | ERF2 | 84 | 0 | 0 | 0 | 0 | 0 | 0 | 17,27 | 17,27 | Manavella et al. 2006 |
| **cluster721** | **PP2C37 (PP2CA)** | 179 | 11 | 0 | 6 | 1 | 2 | 0 | 13,17,19,35 | 6,13,17,19,27,35 | Mishra et al. 2006 |
| cluster7414 | ERF5 | 86 | 9 | 0 | 1 | 2 | 3 | 0 | 17,27,30 | 17,19,27,30 | Pan et al. 2012 |
| cluster8059 | WIN1 (SHN1) | 79 | 4 | 0 | 0 | 0 | 0 | 1 | 17,27 | 0,17,27 | Sajeevan et al. 2017 |
| cluster837 | CRPK1 | 177 | 5 | 0 | 3 | 2 | 0 | 0 | 13,20,27 | 0,13,20,27,35 | Liu et al. 2017 |
| cluster910 | HSP70 | 221 | 10 | 2 | 2 | 0 | 2 | 0 | 0,27,35 | 0,17,27,35 | Kiang and Tsokos 1998 |

960 NOTE.—TN: tip number; TS: Total lineages under positive selection; Number of lineages under positive selection within M7:

961 Mulloginaceae; Cac35: Cactaceae; Did27: Didiereoideae; Mont: Montiaceae; Port: Portulacaceae. SH80: Duplication nodes with SH-like

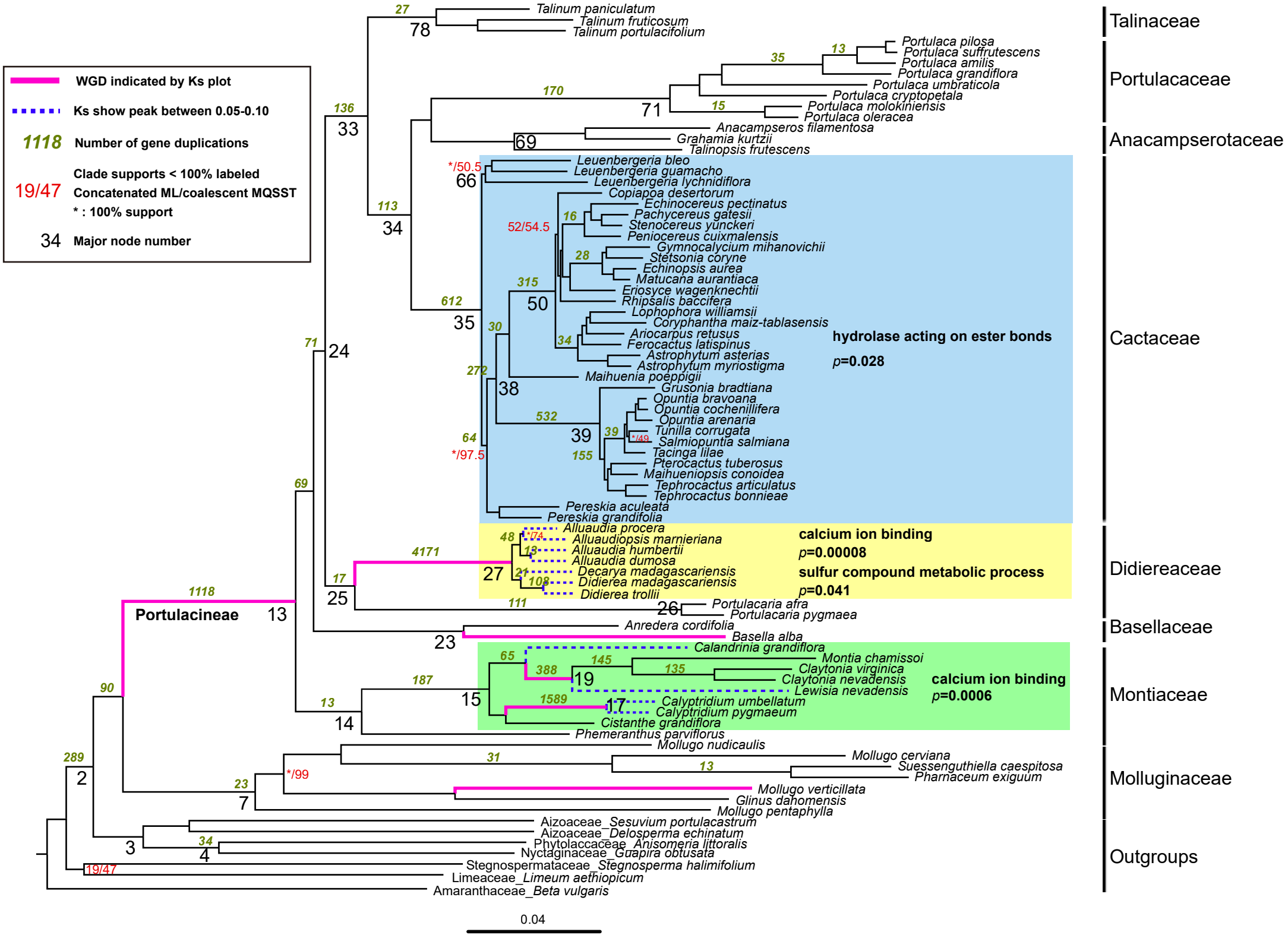
962 support > 80, No support constraint: Duplication nodes with no constraint on support. Node number can be referred to in Fig. 1.

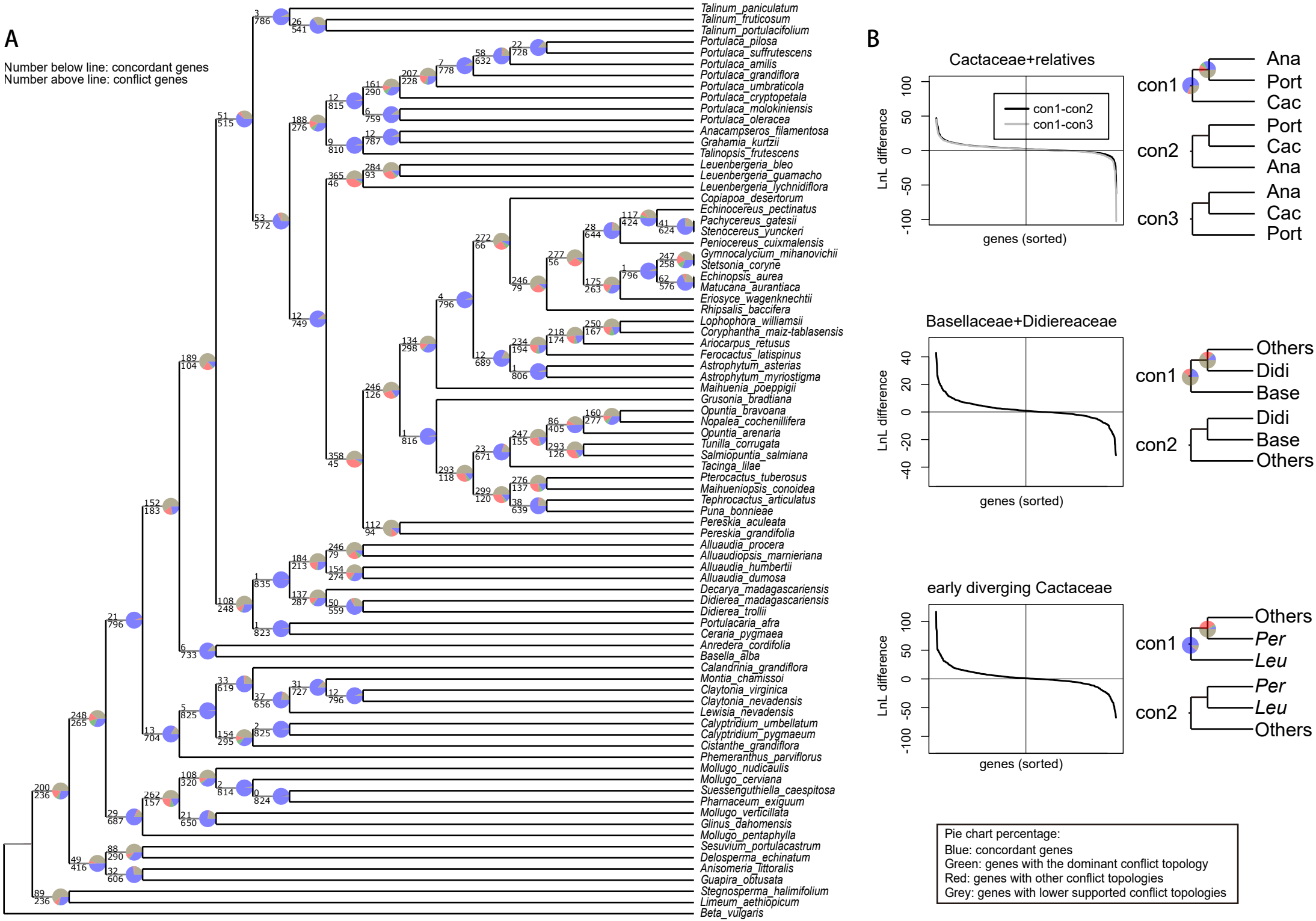
Figure Legends

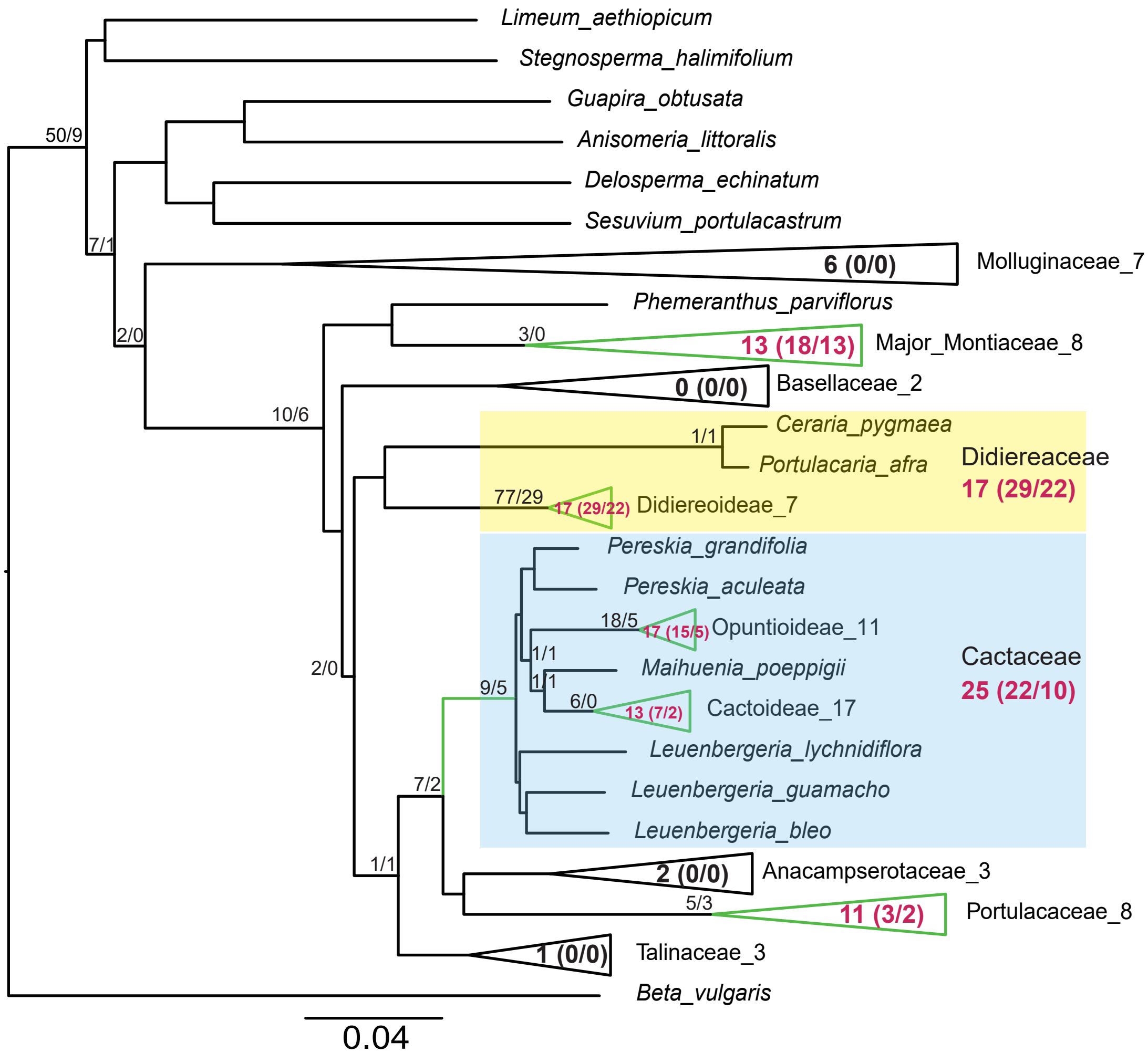
Fig. 1. Species tree from RAxML analysis of the 841-gene supermatrix. Gene duplication numbers (number > 12 are shown) are calculated only for branches with strong support (SH-like > 80) in the 8332 rooted clusters. Some Gene Ontology terms overrepresented in clades Cactaceae, Didiereaceae and Montiaceae are shown with significant level as $p < 0.05$ after Bonferroni correction.

Fig. 2. A: conflict results generated by phyparts. B: Comparison of ln*L* for each individual gene (sorted by greatest differences in likelihood) for alternative phylogenetic hypotheses. Top: Three alternative resolutions were compared involving Cactaceae (Cac), Portulacaceae (Port), and Anacampserotaceae (Ana). Values above the x-axis show the difference in likelihood for those genes supported the resolution found in our result "con1". Those falling below the axis support "con2" (black) or "con3" (gray). Middle: Two alternative hypotheses involving Basellaceae (Base) and Didiereaceae (Didi). Lower: Two alternative hypotheses involving the early diverging Cactaceae (*Per*: *Pereskia*, *Leu*: *Leuenbergeria*). Outlying genes (those genes with extremely high support) are discussed in the Results and Discussion.

Fig. 3. Gene duplication and lineage positive selection patterns on a collapsed species tree from the 29 targeted genes (table 2). The number of sampling taxa in each major clade are labeled after clade name.


WGD indicated by Ks plot
Ks show peak between 0.05-0.10
1118 Number of gene duplications
Clade supports < 100% labeled
19/47 Concatenated ML/coalescent MQSST
* : 100% support
34 Major node number
Talinum paniculatum
Talinum fruticosum
Talinum portulacifolium
Portulaca pilosa
Portulaca suffrutescens
Portulaca amilis
Portulaca grandiflora
Portulaca umbraticola
Portulaca cryptopetala
Portulaca molokiniensis
Portulaca oleracea
Anacampseros filamentosa
Grahamia kurtzii
Talinopsis frutescens
Leuenbergeria bleo
Leuenbergeria guamacho
Leuenbergeria lychnidiflora
Copiapoa desertorum
Echinocereus pectinatus
Pachycereus gatesii
Stenocereus yunckeri
Peniocereus cuixmalensis
Gymnocalycium mihanovichii
Stetsonia coryne
Echinopsis aurea
Matucana aurantiaca
Eriosyce wagenknechtii
Rhipsalis baccifera
Lophophora williamsii
Coryphantha maiz-tablasensis
Ariocarpus retusus
Ferocactus latispinus
Astrophytum asterias
Astrophytum myriostigma
Maihuenia poeppigii
Grusonia bradtiana
Opuntia bravoana
Opuntia cochenillifera
Opuntia arenaria
Tunilla corrugata
Salmiopuntia salmiana
Tacinga lilae
Pterocactus tuberosus
Maihueniopsis conoidea
Tephrocactus articulatus
Tephrocactus bonnieae
Pereskia aculeata
Pereskia grandifolia
hydrolase acting on ester bonds
p=0.028
Alluaudia procera
Alluaudiopsis marnieriana
Alluaudia humbertii
Alluaudia dumosa
Decarya madagascariensis
Didierea madagascariensis
Didierea trollii
calcium ion binding
p=0.00008
sulfur compound metabolic process
p=0.041
Portulacaria afra
Portulacaria pygmaea
Anredera cordifolia
Basella alba
Calandrinia grandiflora
Montia chamissoi
Claytonia virginica
Claytonia nevadensis
Lewisia nevadensis
Calyptridium umbellatum
Calyptridium pygmaeum
Cistanthe grandiflora
Phemeranthus parviflorus
calcium ion binding
p=0.0006
Portulacineae
Mollugo nudicaulis
Mollugo cerviana
Suessenguthiella caespitosa
Pharnaceum exiguum
Mollugo verticillata
Glinus dahomensis
Mollugo pentaphylla
Aizoaceae_Sesuvium portulacastrum
Aizoaceae_Delosperma echinatum
Phytolaccaceae_Anisomeria littoralis
Nyctaginaceae_Guapira obtusata
Stegnospermataceae_Stegnosperma halimifolium
Limeaceae_Limeum aethiopicum
Amaranthaceae_Beta vulgaris
Talinaceae
Portulacaceae
Anacampserotaceae
Cactaceae
Didiereaceae
Basellaceae
Montiaceae
Molluginaceae
Outgroups
0.04


A
Number below line: concordant genes
Number above line: conflict genes
Talinum_paniculatum
Talinum_fruticosum
Talinum_portulacifolium
Portulaca_pilosa
Portulaca_suffrutescens
Portulaca_amilis
Portulaca_grandiflora
Portulaca_umbraticola
Portulaca_cryptopetala
Portulaca_molokiniensis
Portulaca_oleracea
Anacampseros_filamentosa
Grahamia_kurtzii
Talinopsis_frutescens
Leuenbergeria_bleo
Leuenbergeria_guamacho
Leuenbergeria_lychnidiflora
Copiapoa_desertorum
Echinocereus_pectinatus
Pachycereus_gatesii
Stenocereus_yunckeri
Peniocereus_cuixmalensis
Gymnocalycium_mihanovichii
Stetsonia_coryne
Echinopsis_aurea
Matucana_aurantiaca
Eriosyce_wagenknechtii
Rhipsalis_baccifera
Lophophora_williamsii
Coryphantha_maiz-tablasensis
Ariocarpus_retusus
Ferocactus_latispinus
Astrophytum_asterias
Astrophytum_myriostigma
Maihuenia_poeppigii
Grusonia_bradtiana
Opuntia_bravoana
Nopalea_cochenillifera
Opuntia_arenaria
Tunilla_corrugata
Salmiopuntia_salmiana
Tacinga_lilae
Pterocactus_tuberosus
Maihueniopsis_conoidea
Tephrocactus_articulatus
Puna_bonnieae
Pereskia_aculeata
Pereskia_grandifolia
Alluaudia_procera
Alluaudiopsis_marnieriana
Alluaudia_humbertii
Alluaudia_dumosa
Decarya_madagascariensis
Didierea_madagascariensis
Didierea_trollii
Portulacaria_afra
Ceraria_pygmaea
Anredera_cordifolia
Basella_alba
Calandrinia_grandiflora
Montia_chamissoi
Claytonia_virginica
Claytonia_nevadensis
Lewisia_nevadensis
Calyptridium_umbellatum
Calyptridium_pygmaeum
Cistanthe_grandiflora
Phemeranthus_parviflorus
Mollugo_nudicaulis
Mollugo_cerviana
Suessenguthiella_caespitosa
Pharnaceum_exiguum
Mollugo_verticillata
Glinus_dahomensis
Mollugo_pentaphylla
Sesuvium_portulacastrum
Delosperma_echinatum
Anisomeria_littoralis
Guapira_obtusata
Stegnosperma_halimifolium
Limeum_aethiopicum
Beta_vulgaris
B
Cactaceae+relatives
LnL difference
genes (sorted)
con1-con2
con1-con3
con1
con2
con3
Ana
Port
Cac
Port
Cac
Ana
Ana
Cac
Port
Basellaceae+Didiereaceae
LnL difference
genes (sorted)
con1
con2
Others
Didi
Base
Didi
Base
Others
early diverging Cactaceae
LnL difference
genes (sorted)
con1
con2
Others
Per
Leu
Per
Leu
Others
Pie chart percentage:
Blue: concordant genes
Green: genes with the dominant conflict topology
Red: genes with other conflict topologies
Grey: genes with lower supported conflict topologies




**Above branches:**
**Total number of duplications for that node from 29 genes, SH>0/SH>80, resulting from Phyparts analyses**
**21 (27/18):**
**Number of Genes that: have positive selection (have duplications SH>0/SH>80) within the whole clade**